

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-166981

(P2001-166981A)

(43) 公開日 平成13年6月22日 (2001.6.22)

(51) Int.Cl. ⁷	識別記号	F I	テマコード* (参考)
G 0 6 F 12/00	5 4 6	G 0 6 F 12/00	5 4 6 P 5 B 0 7 5
3/00	6 5 1	3/00	6 5 1 A 5 B 0 8 2
13/00	3 5 4	13/00	3 5 4 D 5 B 0 8 9
17/30		15/401	3 2 0 Z 5 E 5 0 1
		15/403	3 7 0 Z
審査請求 未請求 請求項の数14 O L (全 35 頁) 最終頁に続く			

(21) 出願番号 特願平11-345652

(22) 出願日 平成11年12月6日 (1999.12.6)

(71) 出願人 000005496

富士ゼロックス株式会社

東京都港区赤坂二丁目17番22号

(72) 発明者 中山 雄大

神奈川県足柄上郡中井町境430 グリーン

テクなかい 富士ゼロックス株式会社内

(72) 発明者 加藤 裕樹

神奈川県足柄上郡中井町境430 グリーン

テクなかい 富士ゼロックス株式会社内

(74) 代理人 100086531

弁理士 澤田 俊夫

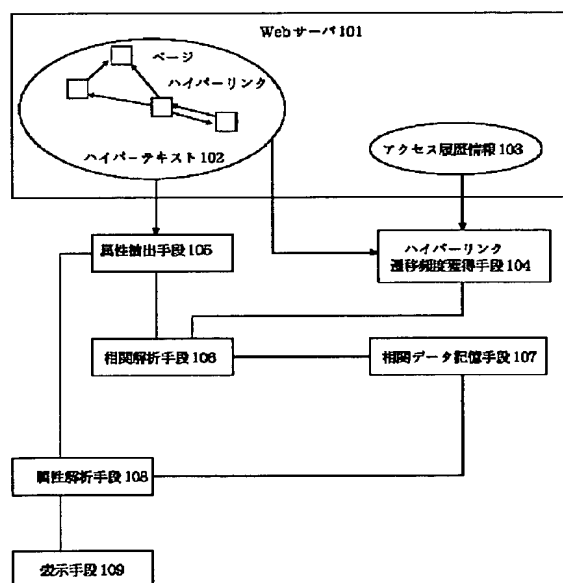
最終頁に続く

(54) 【発明の名称】 ハイパーテキスト解析装置および方法

(57) 【要約】

【課題】 ハイパーテキストシステムのページ構成の問題の原因・改善策を提示し、容易にページ構成を改善できるようにする。

【解決手段】 ハイパーリンク遷移頻度獲得手段104は、ハイパーテキスト102の各ページ組について、アクセス履歴情報103を参照してページ間（つまり、ハイパーリンクの）遷移頻度を算出する。属性抽出手段105は、ハイパーテキスト102の各ページから属性を抽出する。相関解析手段106では、ハイパーリンクの遷移頻度と、属性との相関を計算する。所定のページ組が解析対象として与えられると、属性抽出手段105が、属性を抽出し、この属性が、属性解析手段108において、相関データ記憶手段107において記憶されている重回帰式と比較される。この結果得られた相関式は表示手段109において表示される。



【特許請求の範囲】

【請求項1】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、

ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出手段と、

該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項2】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、

ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出手段と、

該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ群のうちハイパーリンクで繋がれたページ間について、該ハイパーリンク遷移頻度獲得手段を用いてハイパーリンク遷移頻度を計算し、該ハイパーリンク遷移頻度が小さなページ組を獲得する解析対象ページ組獲得手段と、

該解析対象ページ組獲得手段で獲得されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項3】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、

ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出手段と、

該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ群のうちハイパーリンクで繋がれたページ間について、該ハイパーリンク遷移頻度獲得手段を用いてハイパーリンク遷移頻度を計算し、さらに、該属性抽出手段を用いて、該ページ間のコンテンツ類似度を計算し、該ハイパーリンク遷移頻度と該コンテンツ類似度の比率を基に、コンテンツが類似しているにもかかわらずハイパーリンク遷移頻度が小さなページ組を獲得する解析対象ページ組獲得手段と、

該解析対象ページ組獲得手段で獲得されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

20 該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項4】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、

該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出手段と、

該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、

30 該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

40 該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項5】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、

該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出手段と、

50 該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ群の任意のページ間について、該アクセス類似度解析手段を用いてページ間アクセス類似度を計算し、該ページ間アクセス類似度が小さなページ組を獲得する解析対象ページ組獲得手段と、

該解析対象ページ組獲得手段で獲得されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項6】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、

該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出手段と、

該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ群の任意のページ間について、該アクセス類似度解析手段を用いてページ間アクセス類似度を計算し、さらに、該属性抽出手段を用いて、該ページのコンテンツ類似度を計算し、該ページ間アクセス類似度と該コンテンツ類似度の比率を基に、コンテンツが類似しているにもかかわらずページ間アクセス類似度が小さなページ組を獲得する解析対象ページ組獲得手段と、

該解析対象ページ組獲得手段で獲得されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項7】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、

ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出手段と、

該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段と、この表示結果を参照しながら該指定されたページ組のコンテンツを改変する編集手段と、

該編集手段において改変されたコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を予測して改変による効果を計算するコンテンツ改変効果解析手段とを有し、

該表示手段において該コンテンツ改変効果解析手段で得られた結果を表示することを特徴とするハイパーテキスト解析装置。

【請求項8】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、

該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出手段と、

該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段と、この表示結果を参照しながら該指定されたページ組のコンテンツを改変する編集手段と、

該編集手段において改変されたコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のアクセス類似度を予測して改変による効果を計算するコンテンツ改変効果解析手段とを有し、

該表示手段において該コンテンツ改変効果解析手段で得られた結果を表示することを特徴とするハイパーテキスト解析装置。

【請求項 9】 請求項 1 及至請求項 8 に記載のハイパーテキスト解析装置において、相関解析手段が、ハイパーリンク遷移頻度と属性の相関、あるいは、アクセス類似度と属性の相関を計算する際に、有効ではないとおもわれる属性を選択する機能を有し、以下の処理においてこれらの属性を無視することを特徴とするハイパーテキスト解析装置。

【請求項 10】 請求項 1 及至請求項 9 に記載のハイパーテキスト解析装置において、解析対象となるハイパーテキストシステムをネットワーク経由で指定する解析対象指定手段を有することを特徴とするハイパーテキスト解析装置。

【請求項 11】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得ステップと、ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出ステップと、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析ステップと、該相関解析ステップで得られたデータを記憶する相関データ記憶ステップと、指定されたページ組のコンテンツから該属性抽出ステップにより属性を抽出し、これを該相関データ記憶ステップにより記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析ステップと、該属性解析ステップで得られた結果を表示する表示ステップとを有することを特徴とするハイパーテキスト解析方法。

【請求項 12】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析ステップと、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出ステップと、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析ステップと、該相関解析ステップで得られたデータを記憶する相関データ記憶ステップと、指定されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出ステップにより属性を抽出し、これを該相関データ記憶ステップにより記憶されている相関データと比較することによって、該

ページ組間のページ間アクセス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析ステップと、該属性解析ステップで得られた結果を表示する表示ステップとを有することを特徴とするハイパーテキスト解析方法。

【請求項 13】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得ステップと、ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出ステップと、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析ステップと、該相関解析ステップで得られたデータを記憶する相関データ記憶ステップと、指定されたページ組のコンテンツから該属性抽出ステップにより属性を抽出し、これを該相関データ記憶ステップにより記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析ステップと、該属性解析ステップで得られた結果を表示する表示ステップとをコンピュータに実行させるために用いるプログラムを記録したハイパーテキスト解析用のコンピュータ読み取り可能な記録媒体。

【請求項 14】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析ステップと、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出ステップと、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析ステップと、該相関解析ステップで得られたデータを記憶する相関データ記憶ステップと、指定されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出ステップにより属性を抽出し、これを該相関データ記憶ステップにより記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析ステップと、該属性解析ステップで得られた結果を表示する表示ステップとをコンピュータに実行させるために用いるプログラムを記録したハイパーテキスト解析用のコンピュータ

読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ネットワーク上に構成されるハイパーテキストシステムにおいて、その構成に問題がある部分を改善する技術に関する。

【0002】

【従来の技術】ネットワーク上に構成されるハイパーテキストシステム（例えば、World Wide Web；以降Webと略す）では、ハイパーテキストを格納しているサーバにおいてユーザー（訪問者）のアクセス履歴を記録することができる。このアクセス履歴には、一般に、アクセスしてきたユーザーが使用しているコンピュータの識別子（インターネットを利用しているのであればIPアドレス）、アクセスしてきた時刻、アクセスしたページ（＝ファイル）のサーバ上での識別子（WebではURL）が含まれる。

【0003】アクセス履歴とハイパーリンク構造を解析して、ハイパーテキストシステム（例えば、Webサイト）の構成の優劣を判断するような知識を得る技術としては、[Perkowitz and Etzion, 98] (Perkowitz and Etzioni, Adaptive Web Sites: Automatically Synthesizing Web Pages, in Proc. of AAAI-98) がある。

【0004】[Perkowitz and Etzioni, 98] は、まず、サイト上の全てのページ組について、同一ユーザーからのアクセス共起頻度を計算し、予め定められた閾値を超える組を残し、その他を破棄する。さらに、残った組の中から実際にハイパーリンクで連結されている組を破棄する。次に、残った組をアークで連結されたグラフとみなし、各グラフを解析してクリーク（全てのノード間にアークが存在する完全グラフ）を抽出する。このクリークを構成するページ群は、互いに（強い）関連があるがハイパーリンクで連結されていないものなので、サイト上の劣っている部分を発見できたと解釈することができる。この手法では、多くのユーザーが同一セッションでアクセスしてくる傾向が強いにもかかわらずそれらの間にハイパーリンクが存在しない（そのため、ユーザーはページ間遷移に多くの労力を求められるであろう）ページ群を発見できる。

【0005】

【発明が解決しようとする課題】従来技術では、構成の劣ったページ群を明示的に提示することはできるが、該ページ群の問題の原因を解析することはできなかった。そのため、ハイパーテキストシステム管理者は、ページ群の改善方法を試行錯誤によって探らなければならなかった。これは、ある部分を変更してみても、しばらくの間アクセス履歴を収集し、改善効果を測定するという操作

を繰り返すもので、この間、ハイパーテキストシステムの構成は不安定となり、繰り返しアクセスしてくるユーザーを混乱させてしまうという問題が生じる。また、ハイパーテキストシステムにおけるページ群を改善するための手法は、そのハイパーテキストシステムの目的、規模、機能、レイアウト、話題、ユーザータイプ等に依存するので、対象ハイパーテキストシステムの特徴を把握しなければならないという課題もある。

【0006】本発明は上記事情を考慮してなされたもので、対象ハイパーテキストシステム上の任意のページ組について、ページコンテンツから抽出される様々な属性とページ間遷移頻度との相関を予め計算しておき、該ハイパーテキストシステム上で構成に問題がある部分についてどの属性をどのように変更すれば構成の改善になるかを提示することを目的とする。

【0007】

【課題を解決するための手段】本発明の第1の側面によれば、上述の目的を達成するために、ハイパーテキスト解析装置に、ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれた（例えば全ての）ページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから（一つ以上の）属性を抽出する属性抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定された（例えば一つの）ページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させる（一般に、上げることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けている。

【0008】この構成においては、設計者がユーザーの遷移を期待してハイパーリンクを用いて関連付けたページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができる。ハイパーリンクシステム例えばWebシステムの管理者は、この提示内容に基づいてWebページの属性を変化させて所望のWebシステムを構築維持することができる。

【0009】また、本発明の第2の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれた（例えば全ての）ページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから（一つ以上の）属性を抽出する属性

抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ群のうちハイパーリンクで繋がれた（例えば全ての）ページ間について、該ハイパーリンク遷移頻度獲得手段を用いてハイパーリンク遷移頻度を計算し、該ハイパーリンク遷移頻度が小さなページ組（例えば一つ以上）を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させる（一般に、上げることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けるようにしている。

【0010】この構成においても、ハイパーリンクを用いて関連付けたページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。さらに、ハイパーリンクでリンクされたページでありながら遷移頻度の少ないページを修正対象として自動的に選別することができる。

【0011】また、本発明の第3の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれた（例えば全ての）ページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから（一つ以上の）属性を抽出する属性抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ群のうちハイパーリンクで繋がれた（例えば全ての）ページ間について、該ハイパーリンク遷移頻度獲得手段を用いてハイパーリンク遷移頻度を計算し、さらに、該属性抽出手段を用いて、該ページ間のコンテンツ類似度を計算し、該ハイパーリンク遷移頻度と該コンテンツ類似度との比率を基に、コンテンツが類似しているにもかかわらずハイパーリンク遷移頻度が小さなページ組（例えば一つ以上）を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させる（一般に、上げることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかとい

う情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けるようにしている。

【0012】この構成においても、ハイパーリンクを用いて関連付けたページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。さらに、ハイパーリンクでリンクされたページであり、かつ関連したコンテンツを有しながら遷移頻度の少ないページを修正対象として自動的に選別することができる。

【0013】また、本発明の第4の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムを構成する（例えば全ての）ページ組について、アクセス履歴情報を解析することによって、（多くの）ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから（一つ以上の）属性を抽出する属性抽出手段と、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定された（例えば一つの）ページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させる（一般に、大きくすることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けるようにしている。

【0014】この構成においては、リンクにより直接繋がっていないページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。

【0015】また、本発明の第5の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムを構成する（例えば全ての）ページ組について、アクセス履歴情報を解析することによって、（多くの）ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから（一つ以上の）属性を抽出する属性抽出手段と、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ群の任意のページ間について、該アクセス類似度解析手段を用いてページ間アクセス類似度を計算し、該ページ間

アクセス類似度が小さなページ組（一つ以上）を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させる（一般に、大きくすることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けるようにしている。

【0016】この構成においても、リンクにより直接繋がっていないページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。さらに、ユーザによるアクセスが共起することの少ないページ間を修正対象として自動的に選別することができる。

【0017】また、本発明の第6の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムを構成する（例えば全ての）ページ組について、アクセス履歴情報を解析することによって、（多くの）ユーザが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから（一つ以上の）属性を抽出する属性抽出手段と、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ群の任意のページ間について、該アクセス類似度解析手段を用いてページ間アクセス類似度を計算し、さらに、該属性抽出手段を用いて、該ページ間のコンテンツ類似度を計算し、該ページ間アクセス類似度と該コンテンツ類似度の比率を基に、コンテンツが類似しているにもかかわらずページ間アクセス類似度が小さなページ組（一つ以上）を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させる（一般に、大きくすることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けるようにしている。

【0018】この構成においても、リンクにより直接繋がっていないページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することがで

き、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。さらに、コンテンツが関連していながらユーザによるアクセスが共起することの少ないページ間を修正対象として自動的に選別することができる。

【0019】また、本発明の第1～第6の側面において、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるハイパーリンクの位置を抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるハイパーリンクの数を抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるハイパーリンクの種類を抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるハイパーリンクのサイズを抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるハイパーリンクを表わす文字の種類を抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるページのサイズを抽出するようにしてもよい。

【0020】また、本発明の第1、第2、第4または第5の側面において、属性抽出手段が、属性の一つとして、少なくとも、ページ組間のコンテンツ類似度を抽出するようにしてもよい。

【0021】また、本発明の第1～第6の側面において、属性抽出手段が、属性の一つとして、少なくとも、ページの位置を抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページの更新日時を抽出するようにしてもよい。

【0022】また、本発明の第4～第6の側面において、属性抽出手段が、属性の一つとして、少なくとも、ページ間のハイパーリンク遷移数を抽出するようにしてもよい。

【0023】また、本発明の第1～第3の側面において、ハイパーリンク遷移頻度解析手段が、ハイパーテキストを網羅的にアクセスして自動的に情報を収集する探索ロボットを識別して、当該探索ロボットからのアクセス情報を排除して解析を行うようにしてもよい。

【0024】また、本発明の第4～第6の側面において、アクセス類似度解析手段が、ハイパーテキストを網羅的にアクセスして自動的に情報を収集する探索ロボットを識別して、当該探索ロボットからのアクセス情報を排除して解析を行うようにしてもよい。

【0025】また、本発明の第1～第3の側面において、ハイパーリンク遷移頻度解析手段が、ブロックシーサーバを識別して、当該ブロックシーサーバからのアクセス情報を排除して解析を行うようにしてもよい。

【0026】また、本発明の第4～第6の側面におい

て、アクセス類似度解析手段が、ブロッカーサーバーを識別して、当該ブロッカーサーバーからのアクセス情報を排除して解析を行うようにしてもよい。

【0027】また、本発明の第1～第3の側面において、ハイパーリンク遷移頻度解析手段が、被リンク数が大きなページをリンク先とするハイパーリンク遷移を排除して解析を行うようにしてもよい。

【0028】また、本発明の第7の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造を解析して、ハイパーリンクで繋がれた（例えば全ての）ページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから（一つ以上の）属性を抽出する属性抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定された（例えば一つの）ページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させる（一般に、上げることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段と、この表示結果を参照しながら該指定されたページ組のコンテンツを変更する編集手段と、該編集手段において改変されたコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を予測して改変による効果を計算するコンテンツ改変効果解析手段とを設け、該表示手段において該コンテンツ改変効果解析手段で得られた結果を表示するようにしている。

【0029】この構成においては、ハイパーリンクを用いて関連付けたページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、また属性を調整した後の予想結果に基づいて確実に属性調整を行え、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。

【0030】また、本発明の第8の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムを構成する（例えば全ての）ページ組について、アクセス履歴情報を解析することによって、（多くの）ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから（一つ以上の）属性を抽出する属性抽出手段と、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータ

を記憶する相関データ記憶手段と、指定された（例えば一つの）ページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させる（一般に、大きくすることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段と、この表示結果を参照しながら該指定されたページ組のコンテンツを改変する編集手段と、該編集手段において改変されたコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のアクセス類似度を予測して改変による効果を計算するコンテンツ改変効果解析手段とを設け、該表示手段において該コンテンツ改変効果解析手段で得られた結果を表示するようにしている。

【0031】この構成においては、リンクにより直接繋がっていないページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、また属性を調整した後の予想結果に基づいて確実に属性調整を行え、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。

【0032】また、本発明の第1～第6の側面において、相関解析手段が、ハイパーリンク遷移頻度と属性との間の相関（あるいは、アクセス類似度と属性との間の相関）を計算する際に、有効ではないとおもわれる属性を選択する機能を有し、後続する処理においてこれらの属性を無視するようにしてもよい。

【0033】また、本発明の第1～第6の側面において、解析対象となるハイパーテキストシステムを（例えばネットワーク経由で）指定する解析対象指定手段を設けるようにしてもよい。

【0034】また、本発明の第1～第6の側面において、複数のユーザーが、解析対象となるハイパーテキストシステムを（例えばネットワーク経由で）指定する解析対象指定手段を設けるようにしてもよい。

【0035】また、本発明の第1～第6の側面において、解析対象となるハイパーテキストシステムのコンテンツを（例えばネットワーク経由で）受け渡すコンテンツ伝送手段を設けるようにしてもよい。

【0036】また、本発明の第1～第6の側面において、複数のユーザーが、解析対象となるハイパーテキストシステムのコンテンツを（例えばネットワーク経由で）受け渡すコンテンツ伝送手段を有するようにしてもよい。

【0037】また、本発明の第1～第6の側面において、解析対象となるハイパーテキストシステムのアクセ

ス履歴情報を（例えばネットワーク経由で）受け渡すアクセス履歴情報伝送手段を設けるようにしてもよい。

【0038】また、本発明の第1～第6の側面において、複数のユーザーが、解析対象となるハイパーテキストシステムのアクセス履歴情報を（例えばネットワーク経由で）受け渡すアクセス履歴情報伝送手段を有するようにしてもよい。

【0039】また、本発明の第1～第6の側面において、属性抽出手段が抽出する属性の組を（例えばネットワーク経由で）指定する属性指定手段を設けるようにしてもよい。

【0040】また、本発明の第1～第6の側面において、複数のユーザーが、属性抽出手段が抽出する属性の組を（ネットワーク経由で）指定する属性指定手段を有するようにしてもよい。

【0041】また、本発明の第7の側面において、編集手段がネットワーク経由で操作できるようにしてもよい。また、編集手段がネットワーク経由で複数のユーザーによって操作できるようにしてもよい。

【0042】また、本発明の第1～第6の側面において、表示手段が表示内容をネットワーク経由で伝送できるようにしてもよい。また、表示手段が表示内容をネットワーク経由で複数のユーザーに対して伝送できるようにしてもよい。

【0043】なお、本発明は方法の態様で実現することもできる。この場合、ハイパーリンク解析方法は、ハイパーリンク解析装置を構成する各手段に対応する手順で構成される。また、本発明の構成のうちの少なくともその一部をコンピュータソフトウェアとして実現することも可能である。また本発明を当該コンピュータソフトウェアを記録した記録媒体により実現することも可能である。

【0044】

【発明の実施の形態】ネットワーク上に構成されるハイパーテキストシステムのなかで代表的なものはWebである。以下、HTML（ハイパーテキストマークアップランゲージ）で記述されたWebを例として、本発明の実施例を説明する。

【0045】【実施例1】図1は、本発明の実施例1を示す構成図である。同図において、Webサーバ101はネットワーク上で情報を発信する手段であり、Webサーバ101にはユーザーに提供したい情報が、ページ（＝ファイル）とハイパーリンクからなるハイパーテキスト102として貯えられている。ユーザーはWebサーバ101上のハイパーテキスト102にアクセスすることで情報を入手できる。Webサーバ101は、ユーザーからのアクセスがある毎に、ユーザーのコンピュータを識別するためのコンピュータ識別子（IPアドレス）とアクセス時刻とユーザーのアクセスしたページのあるアドレス（URL）をアクセス履歴情報103とし

て記録する手段を持つ。

【0046】ハイパーリンク遷移頻度獲得手段104は、ハイパーテキスト102の各ページ（起点ページと呼ぶ）について、そこからのリンク先ページ（Webサーバ101上にあるものだけを対象とし、Webサーバ101の外にあるものは無視する）を獲得して、それぞれのページ組について、アクセス履歴情報103を参照して、同一ユーザーから同一セッションで起点ページからリンク先ページの順にアクセスがあれば一件として順次カウントしていく（ある定められた時間枠で、各IPアドレス毎にアクセス先ページ（URL）を時間順に並べることによって容易に実現できる）。得られた積算カウントがページ間（つまり、ハイパーリンクの）遷移頻度である。ここで、本発明の変形例として、積算カウントを起点ページへのアクセス数で割った値をハイパーリンク遷移頻度として扱ってもよい。また別の変形例として、ハイパーリンク遷移頻度獲得手段104が、情報収集のための探索ロボットからのアクセスを無視するようにしてもよい。これは、例えば、ある慣習に従って探索ロボットがアクセスする特殊なページ（Webでは、ルート直下に置かれるrobots.txtという名のファイル）へのアクセスの有無や、短期間に多数のノードを網羅的にアクセスするという探索ロボットに特徴的な振る舞いの有無や、既知の探索ロボットのコンピュータ識別子であるか否かによって、探索ロボットを識別して、該探索ロボットのコンピュータ識別子に関わる情報を排除することによって実現できる。こうすることによって、ハイパーリンク遷移頻度がより正確に実ユーザー（＝人間）の振る舞いを反映したものとなり、本ハイパーテキスト解析装置の精度が向上する。さらに別の変形例として、ハイパーリンク遷移頻度獲得手段104が、プロキシサーバからのアクセスを無視するようにしてもよい。これは、アクセス履歴情報103におけるコンピュータ識別子のうち、プロキシサーバに対応するものは（一般に）複数のユーザーからのアクセスを集約したものであるため、そのハイパーリンク遷移頻度を求めても個別のユーザーの振る舞いを正確に反映したものにはならず解析精度が下がってしまうという問題を解決するために行うものである。プロキシサーバであるか否かは、ある期間に、予め定めておいた閾値（一個人が通常アクセスすることがないようなアクセス頻度を閾値とする）を越える数のページにアクセスしたかどうかや、既知のプロキシサーバのコンピュータ識別子であるか否かによって識別できる。さらに別の変形例として、ハイパーリンク遷移頻度獲得手段104が、ハイパーテキスト102のハイパーリンク構造を解析して各ページの被リンク数を求め、その値が予め定められた閾値を越えるようなページへのハイパーリンク遷移を無視するようにしてもよい。リンク数または被リンク数が大きなページはサイトナビゲート用のインデックスページであることが

多い。これらのページはサイトのコンテンツ提供者にとってユーザーにアクセスしてもらいたいコンテンツを包含するページというよりも、サイトの性能を維持するための機能的役割を果たすためのものである傾向が強く、これらをリンク先とするハイパーリンク遷移頻度を求めても、コンテンツを探索するユーザーの振る舞いを直接的に定量化することにはつながらない。そこで、これらのページをリンク先とするハイパーリンク遷移を無視することによって、解析精度が上がり、さらに計算量も減少して処理効率が良くなる。

【0047】属性抽出手段105は、ハイパーテキスト102の各ページから（一つ以上の）属性を抽出する。属性としては、例えば、以下の9項目の中から任意のN個の組合わせ（Nは1以上9以下）を抽出する。もちろん他の属性を用いてもよい。

【0048】〔属性1〕ページのコンテンツにおける、それぞれのリンク先ページに対応するハイパーリンクの位置を抽出する。これは、ハイパーリンクが埋め込まれている位置がページのコンテンツの先頭から何文字目か（文字カウントでわかる）、あるいは、何単語目か（既存の形態素解析技術を用いることによって実現できる）、あるいは、何パラグラフ目か（タグ情報を用いることによって実現できる）という情報を得ることによって実現できる。また、本発明の変形例として、ハイパーリンクの位置を、ページがブラウザでレンダリングされ、表示される際のウインドウ枠内の位置として定めてもよい。これは、標準となるウインドウ枠サイズを予め定めておき、ハイパーリンクがウインドウ枠内に表示されるかどうか、あるいは、ウインドウ枠を例えば縦に均等3分割して（上部、中部、下部、見えない部分）と領域を定め、ハイパーリンクがどの領域に属するかを抽出する。また、別の変形例として、ハイパーリンクの位置を、ページコンテンツの全てのハイパーリンクの中で何番目に現れるかという順序で定めてもよい。

【0049】〔属性2〕ページのコンテンツにおけるハイパーリンクの数を抽出する。また、本発明の変形例として、特定のページ（＝注目しているリンク先ページ）を指すハイパーリンクの数としてもよい。これは、ハイパーテキスト102の全てのページからハイパーリンクを抽出してそのリンク先URLを調べることによって容易にカウントできる。

【0050】〔属性3〕ページのコンテンツにおけるハイパーリンクがテキストで示されているか、あるいは、画像で示されているかを抽出する。これは、ページのコンテンツからハイパーリンクを表わすタグを取り出し（...）、この中に画像ファイルが埋め込まれている（）か否かを調べることによって実現できる。

【0051】〔属性4〕ページのコンテンツにおけるハイパーリンクのサイズを抽出する。サイズの尺度としては、例えば、テキストであれば文字数や単語数（既存の形態素解析技術を用いることによって実現できる）、画像であれば表示領域の面積（IMGタグ中のwidthとheightの値を掛け合わせて求める）を用いる。

【0052】〔属性5〕ページのコンテンツにおけるハイパーリンクテキストの文字種を抽出する。文字種としては、例えば、フォントサイズ、あるいは、フォント色、イタリックやボールド等の文字飾りの有無を抽出する。これらは、タグによって規定されているので容易に抽出できる。

【0053】〔属性6〕ページのサイズを抽出する。サイズの尺度としては、例えば、コンテンツの文字数や単語数（既存の形態素解析技術を用いることによって実現できる）、あるいはページファイルのサイズ（Byte数）を用いる。

【0054】〔属性7〕起点ページとリンク先ページのコンテンツ類似度を抽出する。コンテンツ類似度の計算法は、既存技術を用いる。例えばベクトル空間モデル法〔Salton and Allan, 1994〕（Salton and Allan, Text Retrieval Using the Vector Processing Model, in Proc. of SDAIR-94.）では、各ページについて、単語を項としその出現頻度を項の値とするようなベクトルを生成して、ベクトル間の内積値をコンテンツ類似度として用いている。ここで、本発明の変形として、各起点ページのハイパーリンクが埋め込まれている位置の近傍テキスト（例えば、同一パラグラフと前後N単語を近傍とする）に出現する単語だけを対象とする（リンク先ページは、ページ全体に出現する単語を対象とする、あるいは、ハイパーリンクにリンク先のページ中の位置まで記述してあれば、その近傍テキストを対象とする）ようにしてもよい。

【0055】〔属性8〕ページ間の論理的な位置差分を抽出する。例えば、Webサーバ101上でページのファイルが置かれているディレクトリの位置を論理的な位置として、ルートからの深さを抽出する。次に、起点ページとリンク先ページの深さの差を計算してその値を論理的な位置差分とする（ルートからの深さは、URLを解析すればわかる。例えば、起点ページがPage1.htmlという名前のファイルで、www.○○.co.jp/Dir1/Page1.htmlというURLで表わされていて、リンク先ページがPage2.htmlという名前のファイルで、www.○○.co.jp/Dir1/Dir2/Page2.htmlというURLで表わされている場合、論理的な位置差分は1である）。

【0056】〔属性9〕ページのファイルの更新日時を調べ（更新日時は、Webサーバ101においてファイ

ルプロパティを調べれば容易にわかる)、本発明のハイパーテキスト解析装置使用時と更新日時の差分時間を属性として抽出する。

【0057】図2は、各ページについて、ハイパーリンク遷移頻度獲得104で得られたハイパーリンク遷移頻度と、属性抽出手段105で得られた属性を示した表の例の一部である。この例では、ハイパーリンク遷移頻度として、積算カウントを起点ページへのアクセス数で割った値を用い、属性の組合わせとして、上記の1(先頭からの単語数カウントを使用)、4(文字数を使用)、

6(起点ページの単語数を使用)の3つを用いている。【0058】図1の相関解析手段106では、ハイパーリンク遷移頻度獲得手段104で得られたハイパーリンク遷移頻度と、属性抽出手段105で得られた属性の相関を計算する。相関の計算法としては、既存の統計手法を用いることができるが、ここでは、重回帰分析を用いた場合の実施例を説明する。本発明に重回帰分析を適用した場合、目的変量はハイパーリンク遷移頻度(y とおく)で、説明変量は各属性($x_1, x_2, x_3, \dots, x_n$ とおく)となる(図2の例では、目的変量は遷移頻度(y)で、説明変量はリンク位置(x_1)とリンクサイズ(x_2)とページサイズ(x_3)の3つとなる)。重回帰分析では、 $y = a_1 x_1 + a_2 x_2 + a_3 x_3 + \dots + a_n x_n + a_0$ (a は回帰係数)という重回帰式を作り、左辺と右辺の y と x にデータを代入して、その差の二乗を表わす式を得る。一般にデータ系列は複数(多数)あるので、それぞれについて左辺と右辺の差の二乗を表わす式を得て、全ての式の和(Q とおく)を得る。次に、 Q を最小にするような回帰係数を求めるという手続きを取る(Q をそれぞれの回帰係数で偏微分してその値を0とおいた等式を作り、その連立方程式を解けばよい)。こうして得られた重回帰式は、図1の相関データ記憶手段107において記憶される。ここで、本発明の変形例として、説明変量の候補の中から、目的変量(=ハイパーリンク遷移頻度)の予測に有効ではないと思われる属性を選択し、以下の処理においてこれらの属性を無視することによって、解析精度を上げ、さらに処理効率を上げるようにしてもよい。重回帰分析において、有効な説明変量の組合わせを選択する手法は既存であり、例えば、forward selection methodやstepwise methodが提案されている(Carl J. Huberty, Applied Discriminant Analysis, John Wiley & Sons, Inc, 1994)ので、本発明ではこれらのうちいずれかを用いれば容易に実現できる。

【0059】次に、あるページ組が解析対象として与えられると、本ハイパーリンク解析装置は、属性抽出手段105を用いて、前述の方法により、属性を抽出する(上で、説明変量として採用した属性だけを抽出すれば

よい。図2の例では、上記の1、4、6の3つの属性を抽出する)。このようにして得られた属性は、属性解析手段108において、相関データ記憶手段107において記憶されている重回帰式と比較される。その方法は、まず、ある一つの属性に注目して(例えば、図2の例で、上記の1の属性に注目したとする)、他の属性の値を重回帰式に代入する(上記の4と6の属性の値をそれぞれ x_1 と x_2 に代入する)。すると、重回帰式は、一変数の一次式で表わされ($y = a_1 x_1 + c$ (c は定数)という形になる)、与えられたページ組のハイパーリンク遷移頻度と注目した属性の(予測)相関式が得られるので、(他の属性は変化させないという前提の基において)該属性をどういう値に変更すれば、ハイパーリンク遷移頻度がどのように変わるかが予測される。尚、ここで、注目している属性の値(属性抽出手段105で得られたもの)を代入すると、現状のハイパーリンク遷移頻度予測値が得られる。本発明では、上記の注目している属性というのは、特に指定されるものではなく、全ての属性を順に注目しながら、上記の処理を繰り返す(つまり、属性の数だけ相関式が得られる)。

【0060】このようにして得られた相関式は表示手段109において表示される。図3は、表示の一例を示すものである。ここでは、図2の例の結果表示の一部(上記の1の属性に注目したものを示している)を示している。同図では、与えられたページ組のハイパーリンク遷移頻度を注目した属性(ハイパーリンクの位置)の(一次)関数として表わしている。ここで、 x 軸の値 p は、現状の属性値であり、 $q (= a_1 p + c)$ は、現状のハイパーリンク遷移頻度予測値となる。同図を参照すると、属性値をどれだけ可変させると(ここでは、ハイパーリンクの位置を何単語目に埋め込むかということになる)、どれだけハイパーリンク遷移頻度を可変させられるかが定量的に予測できる(一般に、ハイパーリンク遷移頻度を高くすることを考えるが、ここでは、ハイパーリンク位置を前にするほど、ハイパーリンク遷移頻度が高くなることが予測でき、その効果は回帰係数 a_1 の値で左右される)。

【0061】[実施例2] つぎに本発明の実施例2について説明する。

【0062】図4は、本発明の実施例2を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、解析対象ページ組獲得手段401により、属性解析手段108で解析するページ組を自動的に取捨選択する。この処理の流れは、先ず、指定されたページ集合(例えば、ある特定のサーバ上の全てのページ、とか、ある特定のサーバ上のある特定のディレクトリ以下に存在する全てのページといった指定)のうちハイパーリンクで直接繋がれたページ組について、ハイパーリンク遷移頻度獲得手段104を用いてハイパーリンク遷移頻度を得る。次に、この情

報は解析対象ページ組獲得手段401に渡され、解析対象ページ組獲得手段401において、ハイパーリンク遷移頻度が予め定められた閾値より小さなページ組を獲得して、これらを属性解析手段108に渡す。属性解析手段108では、上記の実施例の方法で、渡されたページ組の解析を（順次）行う。こうすることによって、本ハイパーテキスト解析装置のユーザーが、予め、（ハイパーリンク遷移頻度）に問題がありそうなページ組を探して指定しなければならないという手間を省略できる。

【0063】ここで、本発明の変形例を示す。先ず上記の処理と同様に、指定されたページ集合のうちハイパーリンクで直接繋がれたページ組について、ハイパーリンク遷移頻度獲得手段104を用いてハイパーリンク遷移頻度を得る。さらに、本変形例では、属性抽出手段105において、前述の方法により該ページ組間のコンテンツ類似度を得る。こうして得られた該ページ組間のハイパーリンク遷移頻度とコンテンツ類似度の情報は解析対象ページ組獲得手段401に渡され、解析対象ページ組獲得手段401において、ハイパーリンク遷移頻度をコンテンツ類似度で割った値が予め定められた閾値より小さなページ組を獲得して、これらを属性解析手段108に渡す。ハイパーリンク遷移頻度をコンテンツ類似度で割った値が小さなページ組は、内容が類似しているにも拘わらず、ユーザーがリンクを遷移してくれる傾向が小さいものであり、改善の余地があるといえる（改善のための指針は上記と同様に属性解析手段108において得られ、表示手段109において表示される）。

【0064】〔実施例3〕つぎに本発明の実施例3について説明する。

【0065】図5は、本発明の実施例3を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。同図において、アクセス類似度解析手段501は、ハイパーテキスト102の（全ての）ページ組について、そのページ間のアクセス類似度を求める。ここで、アクセス類似度とは、与えられた2つのページについて、いかに多くのユーザーが双方をアクセスしたかという度合いを表わす指標であり、例えば、ベクトル空間モデル法を使って、各ページについて、アクセスのあったIPアドレスを項としその出現頻度を項の値とするようなベクトルを生成して、ベクトル間の内積値をアクセス類似度として用いる。このように前述のハイパーリンク遷移頻度の代わりに、アクセス類似度を用いることによって、（2つのページ間のハイパーリンクの有無を考慮することなく）任意のページ組を対象として、そのページ構成の改善策を提示できるようになる。また本発明の変形例としてアクセス類似度解析手段501が、上記実施例と同様の方法で、情報収集のための探索ロボットからのアクセスを無視するようにしてもよい。さらに別の変形例として、アクセス類似度解析手段501が、上記実施例と同様の方法で、ブロッ

ーサーバからのアクセスを無視するようにしてもよい。

【0066】属性抽出手段105は、上記実施例と同様の処理を行うが、ここでは、さらに、ページ間のハイパーリンク遷移数を計算してこれを属性の1つとしてもよい。具体的には、ページ組（ページA、ページB）が与えられたときに、ハイパーリンクの向きを考慮しながらページAからページBへの最短経路の遷移数を計算し、次にページBからページAへの最短経路の遷移数を計算し、両者を比べて小さい方をページ組（ページA、ページB）のハイパーリンク遷移数とする。ここで、ページAからページBへの最短経路は、ページAからページBに到達するまで幅優先で深さ（＝遷移数）をカウントしながらハイパーリンクを探索すればよい。この探索は、一般には、辿ることができるハイパーリンクが無くなるまで続けられるが、処理効率を上げるために、深さ（＝遷移数）が予め決められた閾値に達したところで停止するように設定してもよい。この場合、ページ間の最短経路の遷移数は該閾値とする。

【0067】相関解析手段106では、アクセス類似度解析手段501で得られたアクセス類似度と、属性抽出手段105で得られた属性の相関を上記実施例と同様に重回帰分析によって計算する。得られた重回帰式は、相関データ記憶手段107において記憶される。ここで、本発明の変形例として、上記実施例と同様の方法により、説明変量の候補の中から、目的変量（＝アクセス類似度）の予測に有効ではないとおもわれる属性を選択し、以下の処理においてこれらの属性を無視することによって、解析精度を上げ、さらに処理効率を上げるようにしてもよい。

【0068】次に、あるページ組が解析対象として与えられると、本ハイパーリンク解析装置は、属性抽出手段105を用いて、上記実施例と同様の方法により、属性を抽出する。得られた属性は、属性解析手段108において、相関データ記憶手段107において記憶されている重回帰式と比較され、上記実施例と同様に、与えられたページ組のアクセス類似度と注目した属性の（予測）相関式が得られるので、（他の属性は変化させないという前提の基において）該属性をどういう値に変更すれば、ハイパーリンク遷移頻度がどのように変わるかが予測される。尚、ここで、注目している属性の値（属性抽出手段105で得られたもの）を代入すると、現状のアクセス類似度予測値が得られる。本発明では、上記の注目している属性というのは、特に指定されるものではなく、全ての属性を順に注目しながら、処理を繰り返す（つまり、属性の数だけ相関式が得られる）。このようにして得られた相関式は、上記実施例と同様に、表示手段109において表示される。

【0069】〔実施例4〕つぎに本発明の実施例4について説明する。

【0070】図6は、本発明の実施例4を示す構成図で

ある。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、解析対象ページ組獲得手段601により、属性解析手段108で解析するページ組を自動的に取捨選択する。この処理の流れは、先ず、指定されたページ集合（例えば、ある特定のサーバ上の全てのページ、とか、ある特定のサーバ上のある特定のディレクトリ以下に存在する全てのページといった指定）から任意の2つを取り出したページ組について、アクセス類似度解析手段501を用いてアクセス類似度を得る。次に、この情報は解析対象ページ組獲得手段601に渡され、解析対象ページ組獲得手段601において、アクセス類似度が予め定められた閾値より小さなページ組を獲得して、これらを属性解析手段108に渡す。属性解析手段108では、上記の実施例の方法で、渡されたページ組の解析を（順次）行う。こうすることによって、本ハイパーテキスト解析装置のユーザーが、予め、（アクセス類似度）問題がありそうなページ組を探して指定しなければならないという手間を省略できる。

【0071】ここで、本発明の変形例を示す。先ず上記の処理と同様に、指定されたページ集合から任意の2つを取り出したページ組について、アクセス類似度解析手段501を用いてアクセス類似度を得る。さらに、本変形例では、属性抽出手段105において、前述の方法により該ページ組間のコンテンツ類似度を得る。こうして得られた該ページ組間のアクセス類似度とコンテンツ類似度の情報は解析対象ページ組獲得手段601に渡され、解析対象ページ組獲得手段601において、アクセス類似度をコンテンツ類似度で割った値が予め定められた閾値より小さなページ組を獲得して、これらを属性解析手段108に渡す。アクセス類似度をコンテンツ類似度で割った値が小さなページ組は、内容が類似しているにも拘わらず、ユーザーがリンクを遷移してくれる傾向が小さいものであり、改善の余地があるといえる（改善のための指針は上記と同様に属性解析手段108において得られ、表示手段109において表示される）。

【0072】[実施例5] つぎに本発明の実施例5について説明する。

【0073】図7は、本発明の実施例5を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。尚、本発明では、同図のハイパーリンク遷移頻度獲得手段104をアクセス類似度解析手段501と置き換えることもできる。この場合、以下の説明中のハイパーリンク遷移頻度はアクセス類似度に置き換えられる。

【0074】本実施例は、編集手段701において、本ハイパーテキスト解析装置の操作者が、表示手段109で表示されている情報（指定されたページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果がある

かという情報）を参照しながら指定されたページ組のコンテンツを改変する（＝1つ以上の属性を変化させる）ことができる。図8は、表示手段109における表示の一例である。この例では、P6（起点ページ）とP7（リンク先ページ）が指定されたページ組であり、目的変数 y に対応するハイパーリンク遷移頻度には、積算カウントを起点ページへのアクセス数で割った値を用いている（アクセス類似度を置き換えた場合は、上記実施例と同様に、IPアドレスの出現頻度に基づくベクトルの内積値を用いる）。また、属性の組合わせとして、上記の1（ハイパーリンクの位置：先頭からの単語数カウントを使用）、4（ハイパーリンクサイズ：文字数を使用）、6（ページサイズ：起点ページの単語数を使用）の3つを用いており、これらが順に説明変数 x_1 、 x_2 、 x_3 にそれぞれ対応する。同図の表では、ページ組{P6, P7}から（属性抽出手段105において）抽出された3つの属性値がそれぞれ、 $x_1 = 66$ 、 $x_2 = 7$ 、 $x_3 = 325$ であり、そのとき予測されるハイパーリンク遷移頻度が $y = 0.17$ である（属性解析手段108において求められる）ことがわかる。該操作者は、この $y = 0.17$ というハイパーリンク遷移頻度予測値を増加させたいければ、同図下の3つのグラフを見て、どの属性をどれだけ変化させれば、どれだけ効果があるかということを知ることができる。本例では、 x_1 を小さくする（つまり、P6からP7へのハイパーリンクの位置を前にする）、あるいは、 x_2 を大きくする（つまりP6からP7へのハイパーリンクの文字数を増やす）、あるいは、 x_3 を小さくする（つまりP6の文字数を減らす）ことによってハイパーリンク遷移頻度予測値が増加することがわかる。このとき、各グラフの傾きを見ることによって属性値の可変量がどの程度の効果を持つかも容易に知ることができる（尚、これら3つのグラフはそれぞれに対応する属性だけを変化させたときのハイパーリンク遷移頻度予測値を示すものである）。該操作者は、これらの情報を参考にしながら、編集手段701において、ページ組{P6, P7}のコンテンツを改変できる。このようにして改変されたページ組はハイパーテキスト102に置かれる。

【0075】次に、属性抽出手段105において、上記実施例と同様の方法で、該ページ組の属性が抽出され、コンテンツ改変効果解析手段702に渡される。コンテンツ改変効果解析手段702では、上記実施例と同様の方法で、該ページ組の属性と関連データ記憶手段において記憶されている関連データと比較することによって、該ページ組間のハイパーリンク遷移頻度を予測して改変による効果を計算し、その計算結果を表示手段109に渡す。

【0076】図9は、表示手段109における表示の一例である。同図の表は、ページ組{P6, P7}のコンテンツ改変後のハイパーリンク遷移頻度予測値を示すも

のである。本例では、該操作者は、編集手段701において x_1 と x_2 を変化させており、改変後のハイパーリンク遷移頻度予測値は0.17から0.31に増加している。本発明では、該操作者はこの表示例を参照しながら、編集手段701を用いて、さらに、コンテンツの改変を繰り返すこともできる。このとき、上記実施例と同様に、同図下の3つのグラフを見て、どの属性をどれだけ変化させれば、どれだけ効果があるかということを知ることができる。尚、これら3つのグラフの形状は、図8のコンテンツ改変前のものと比較して、回帰係数

a_1 、 a_2 、 a_3 は同一で、定数項だけが変化したものとなる。

【0077】[実施例6] つぎに本発明の実施例6について説明する。

【0078】図10は、本発明の実施例6を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、解析対象指定手段1001により、本ハイパーテキスト解析装置のユーザーが、解析対象となるWebサーバを指定できるようにする。ユーザーのコンピュータ1002と解析対象指定手段1001はネットワーク（例えばインターネットや電話回線）で接続されている。ユーザーのコンピュータ1002では、解析対象とするWebサーバのURL（サーバを識別する符号であれば何でもよい）が入力され、このURLはネットワークを通じて解析対象指定手段1001に渡される。解析対象指定手段1001は、該URLを属性抽出手段105とハイパーリンク遷移頻度獲得手段104（あるいはアクセス類似度解析手段501）に渡す。属性抽出手段105とハイパーリンク遷移頻度獲得手段104では、HTTP（ハイパーテキストトランスファプロトコル）に従って該URLを発行して、解析対象Webサーバとアクセスして必要な情報を獲得する。ここで、本発明の変形として、解析対象指定手段1001が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、解析対象指定手段1001は、個々のユーザーからユーザーID（あるいはユーザーIDとパスワード）を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。さらに異なる変形として、本発明では、解析対象指定手段1001をWebサーバ上に置き、CGI（コモンゲートウェイインタフェース）スクリプトとして表現して、ユーザーのコンピュータ1002における解析対象指定をWebページのフォームを用いて行えるようにしてもよい。

【0079】[実施例7] つぎに本発明の実施例7について説明する。

【0080】図11は、本発明の実施例7を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、コンテンツ

伝送手段1101により、本ハイパーテキスト解析装置のユーザーが、解析対象となるWebサーバ上のコンテンツを本ハイパーテキスト解析装置に受け渡せるようにする。本実施例では、まず、ユーザーのコンピュータ1002で、解析対象となるWebサーバ上のコンテンツを獲得する（あるいは予め解析対象となるWebサーバ上のコンテンツを持っている）。ユーザーのコンピュータ1002とコンテンツ伝送手段1101はネットワーク（例えばインターネットや電話回線）で接続されており、ユーザーのコンピュータ1002から解析対象であるWebサーバ上のコンテンツがネットワークを通じてコンテンツ伝送手段1101に渡される。コンテンツ伝送手段1101は、該Webサーバ上のコンテンツを属性抽出手段105とハイパーリンク遷移頻度獲得手段104（あるいはアクセス類似度解析手段501）に渡す。Webではコンテンツの改変が頻繁に行われることもあるが、こうすることによって、ユーザーが想定している（時点での）コンテンツを対象とした解析を確実に行うことができる。ここで、本発明の変形として、コンテンツ伝送手段1101が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、コンテンツ伝送手段1101は、個々のユーザーからユーザーID（あるいはユーザーIDとパスワード）を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。

【0081】[実施例8] つぎに本発明の実施例8について説明する。

【0082】図12は、本発明の実施例8を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、アクセス履歴情報伝送手段1201により、本ハイパーテキスト解析装置のユーザーが、解析対象となるWebサーバのアクセス履歴情報を本ハイパーテキスト解析装置に受け渡せるようにする。本実施例では、まず、ユーザーのコンピュータ1002で、解析対象となるWebサーバのアクセス履歴情報を獲得する（あるいは予め解析対象となるWebサーバのアクセス履歴情報を持っている）。ユーザーのコンピュータ1002とアクセス履歴情報伝送手段1201はネットワーク（例えばインターネットや電話回線）で接続されており、ユーザーのコンピュータ1002から解析対象であるWebサーバのアクセス履歴情報がネットワークを通じてアクセス履歴情報伝送手段1201に渡される。アクセス履歴情報伝送手段1201は、該Webサーバのアクセス履歴情報をハイパーリンク遷移頻度獲得手段104（あるいはアクセス類似度解析手段501）に渡す。Webサーバのアクセス履歴情報は個人のプライバシーに関する情報であるとな見なすこともできるため機密性が高い。そのため、Webサーバのアクセス履歴情報が該Webサーバ上でアクセス可能な形態で保持されることは希である。本実施

例のように、ユーザーと本ハイパーテキスト解析装置間で直接、アクセス履歴情報を受け渡すことができれば第3者からアクセス履歴情報にアクセスされる可能性が小さくなる。ここで、本発明の変形として、アクセス履歴情報伝送手段1201が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、アクセス履歴情報伝送手段1201は、個々のユーザーからユーザーID（あるいはユーザーIDとパスワード）を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。

【0083】[実施例9] つぎに本発明の実施例9について説明する。

【0084】図13は、本発明の実施例9を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、属性指定手段1301により、本ハイパーテキスト解析装置のユーザーが、解析対象となるハイパーテキストのコンテンツから抽出する属性の組を指定することできる。ユーザーのコンピュータ1002と属性指定手段1301はネットワーク（例えばインターネットや電話回線）で接続されている。ユーザーのコンピュータ1002では、抽出する属性の組が入力され、この情報はネットワークを通じて属性指定手段1301に渡される。属性指定手段1301は、該情報を属性抽出手段105に渡す。属性抽出手段105では、指定された属性の組を解析対象として処理を行う。ここで、本発明の変形として、属性指定手段1301が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、属性指定手段1301は、個々のユーザーからユーザーID（あるいはユーザーIDとパスワード）を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。さらに異なる変形として、本発明では、属性指定手段1301をWebサーバー上に置き、CGIスクリプトとして表現して、ユーザーのコンピュータ1002における属性の組指定をWebページのフォームを用いて行えるようにしてもよい。この場合、抽出可能な属性を予めメニューとして用意しておき、該フォーム上でユーザーが選択するよな形態にしておいてもよい。

【0085】[実施例10] つぎに本発明の実施例10について説明する。

【0086】図14は、本発明の実施例10を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、ユーザーのコンピュータ1002と編集手段701がネットワーク（例えばインターネットや電話回線）で接続されている。ユーザーのコンピュータ1002における編集作業はネットワークを通じて編集手段701に渡される。ここで、本発明の変形として、編集手段701が複数のユーザーのコンピュータ1002と接続されていてもよ

い。この場合、編集手段701は、個々のユーザーからユーザーID（あるいはユーザーIDとパスワード）を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。さらに異なる変形として、本発明では、編集手段701をWebサーバー上に置き、CGIスクリプトとして表現して、ユーザーのコンピュータ1002における編集作業をWebページのフォームを用いて行えるようにしてもよい。

【0087】[実施例11] つぎに本発明の実施例11について説明する。

【0088】図15は、本発明の実施例11を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、ユーザーのコンピュータ1002と表示手段109がネットワーク（例えばインターネットや電話回線）で接続されている。表示手段109における表示結果はネットワークを通じてユーザーのコンピュータ1002に渡され、ユーザーのコンピュータ1002において表示される。ここで、本発明の変形として、表示手段109が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、表示手段109は、個々のユーザーからユーザーID（あるいはユーザーIDとパスワード）を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。さらに異なる変形として、本発明では、表示手段109をWebサーバー上に置き、表示結果をHTMLフォームで表現して、ユーザーのコンピュータ1002において表示されるようにしてもよい。

【0089】

【発明の効果】以上説明したように、本発明によれば、Webサイト上の任意のページ組について、ページコンテンツから抽出される様々な属性とページ間遷移頻度との相関を基にして、ページ構成の問題の原因を提示し、さらに、その改善策をも提示することができ、ハイパーテキストシステム（Webサイト）管理者は容易にページ構成を改善できる。

【0090】

【図面の簡単な説明】

【0091】

【図1】 本発明の実施例の構成を示すブロック図である。

【0092】

【図2】 本発明の実施例の属性の一例を示す図である。

【0093】

【図3】 本発明の実施例の表示の一例を示す図である。

【0094】

【図4】 本発明の実施例の構成を示すブロック図であ

る。

【0095】

【図5】 本発明の実施例の構成を示すブロック図である。

【0096】

【図6】 本発明の実施例の構成を示すブロック図である。

【0097】

【図7】 本発明の実施例の構成を示すブロック図である。

【0098】

【図8】 本発明の実施例の表示の一例を示す図である。

【0099】

【図9】 本発明の実施例の表示の一例を示す図である。

【0100】

【図10】 本発明の実施例の構成を示すブロック図である。

【0101】

【図11】 本発明の実施例の構成を示すブロック図である。

【0102】

【図12】 本発明の実施例の構成を示すブロック図である。

【0103】

【図13】 本発明の実施例の構成を示すブロック図である。

10

20

*

*【0104】

【図14】 本発明の実施例の構成を示すブロック図である。

【0105】

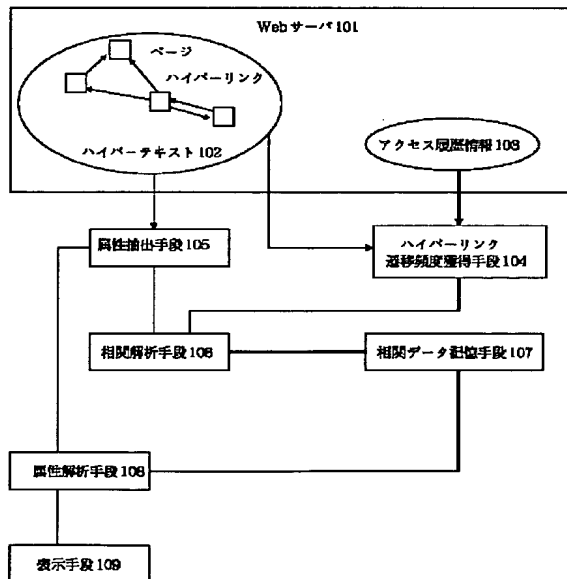
【図15】 本発明の実施例の構成を示すブロック図である。

【0106】

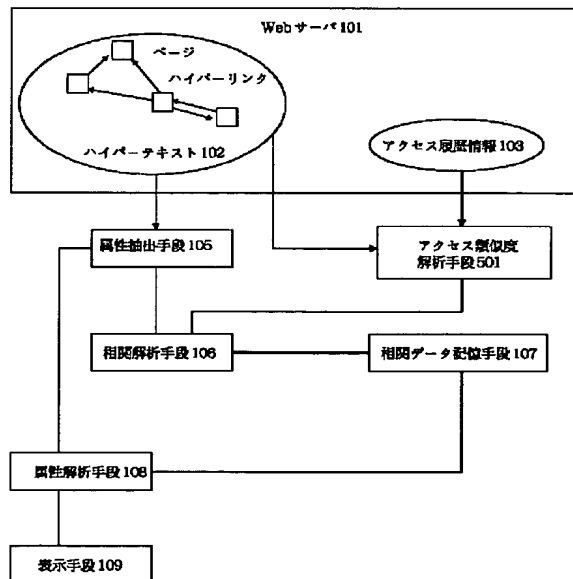
【符号の説明】

101 Webサーバ
102 ハイパーテキスト
103 アクセス履歴情報
104 ハイパーリンク遷移頻度獲得手段
105 属性抽出手段
106 相関解析手段
107 相関データ記憶手段
108 属性解析手段
109 表示手段
401 解析対象ページ組獲得手段
501 アクセス類似度解析手段
601 解析対象ページ組獲得手段
701 編集手段
702 コンテンツ改変効果解析手段
1001 解析対象指定手段
1002 ユーザーのコンピュータ
1101 コンテンツ伝送手段
1201 アクセス履歴情報伝送手段
1301 属性指定手段

【図1】



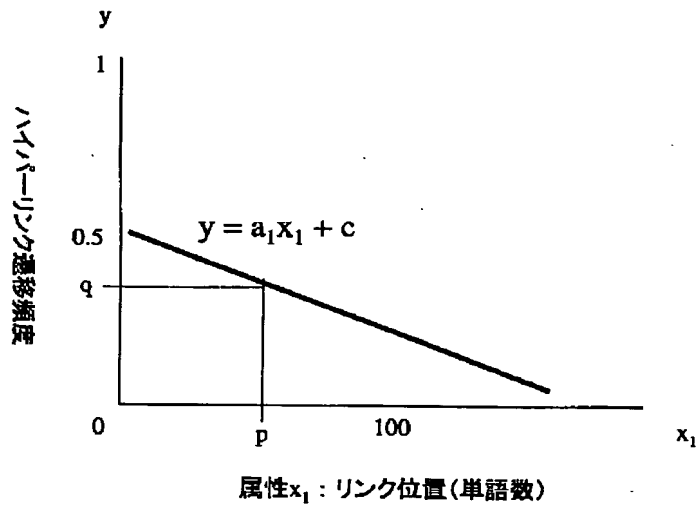
【図5】



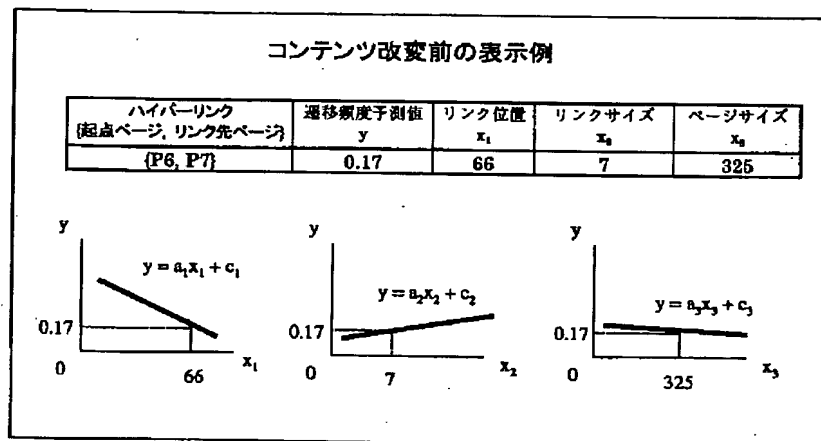
【図2】

ハイパーリンク (起点ページ、リンク先ページ)	遷移頻度	リンク位置	リンクサイズ	ページサイズ
{P1, P2}	0.13	56	8	305
{P1, P3}	0.43	15	7	306
{P1, P4}	0.02	268	3	306
{P2, P4}	0.65	10	25	621
{P2, P5}	0.11	46	2	521
{P3, P2}	0.04	121	18	11
...

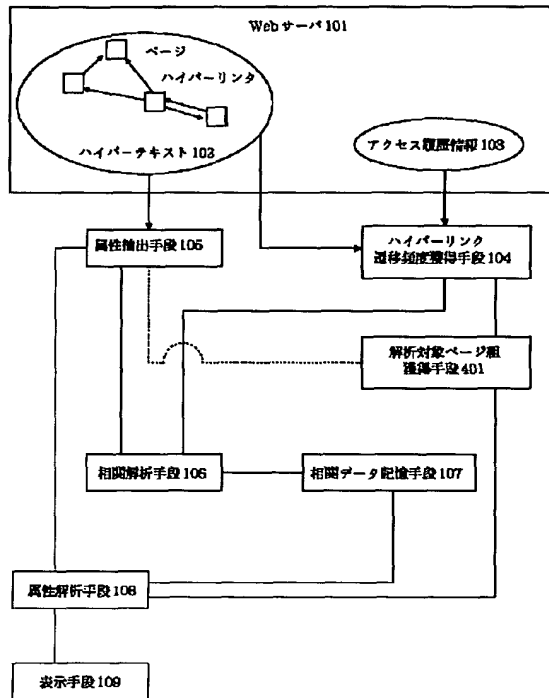
【図3】



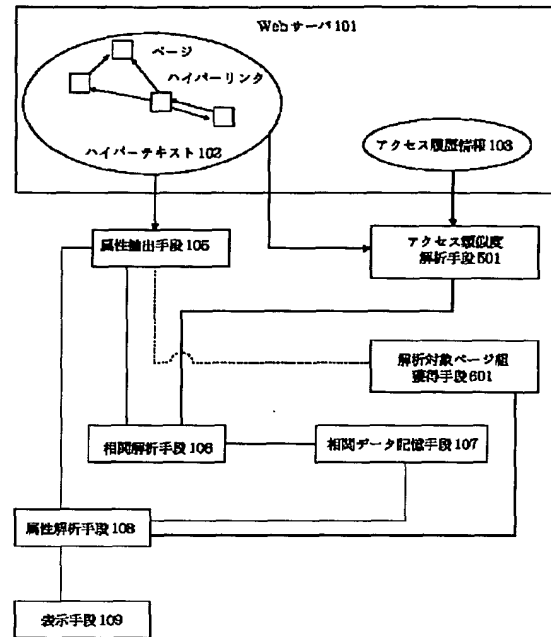
【図8】



【図4】

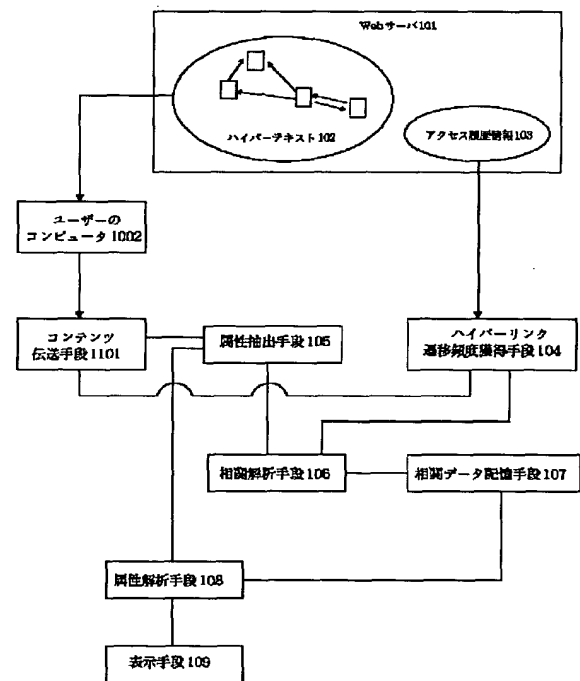
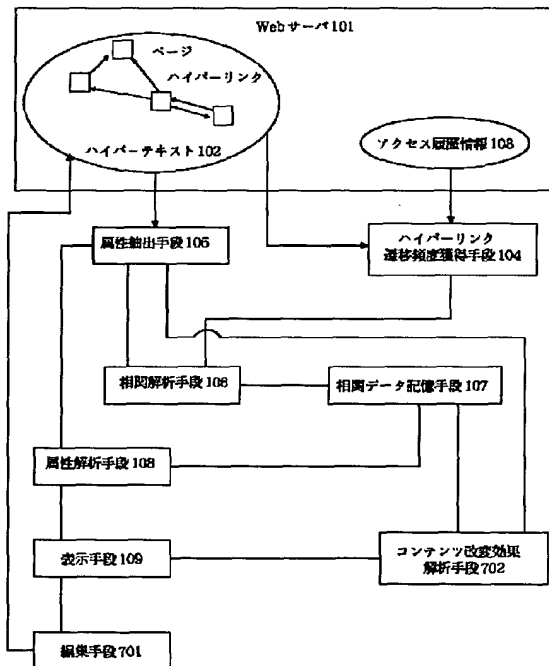


【図6】

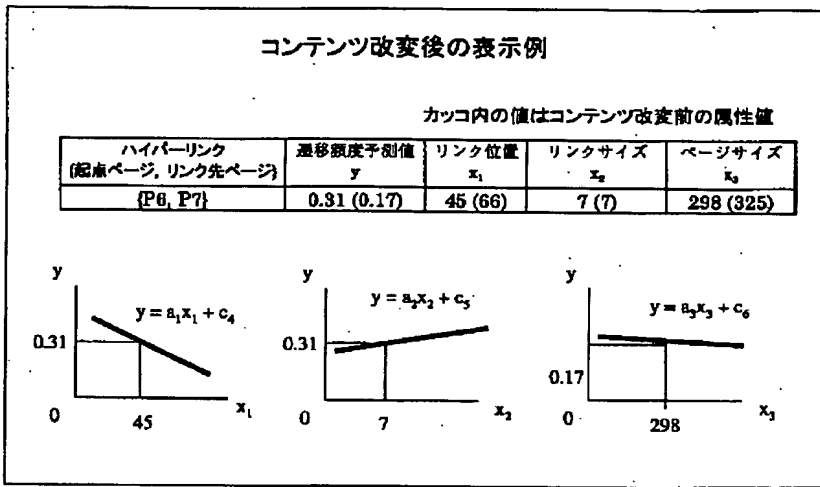


【図11】

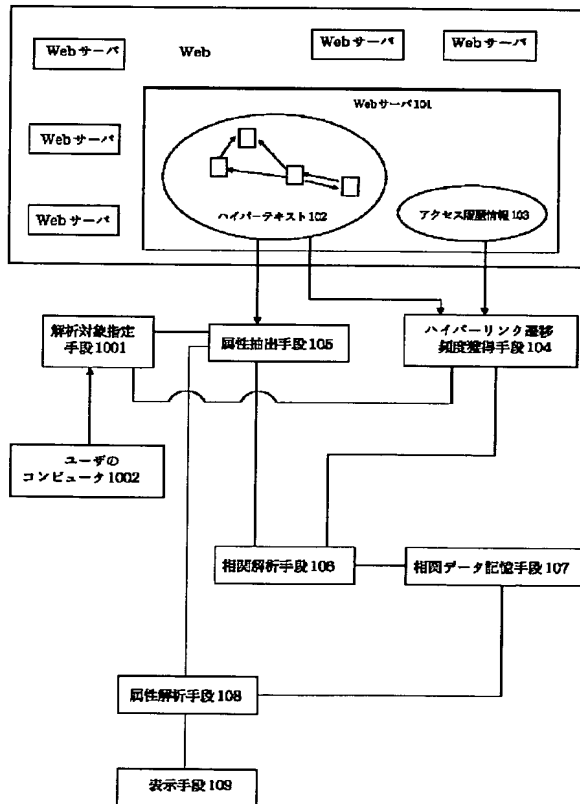
【図7】



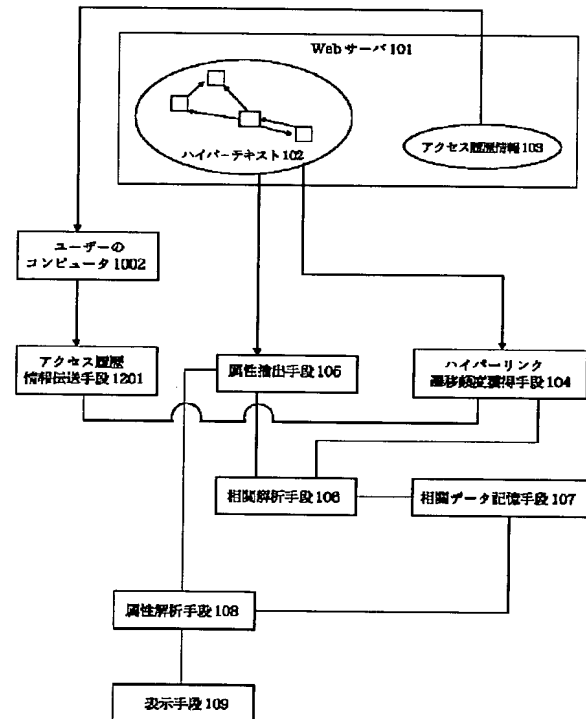
【図9】



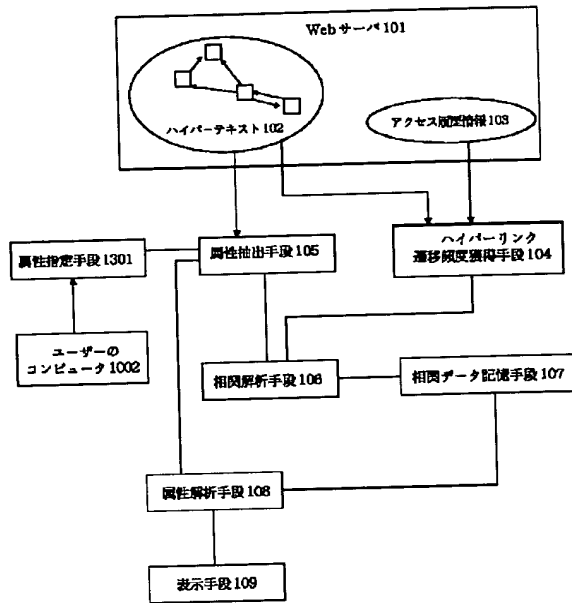
【図10】



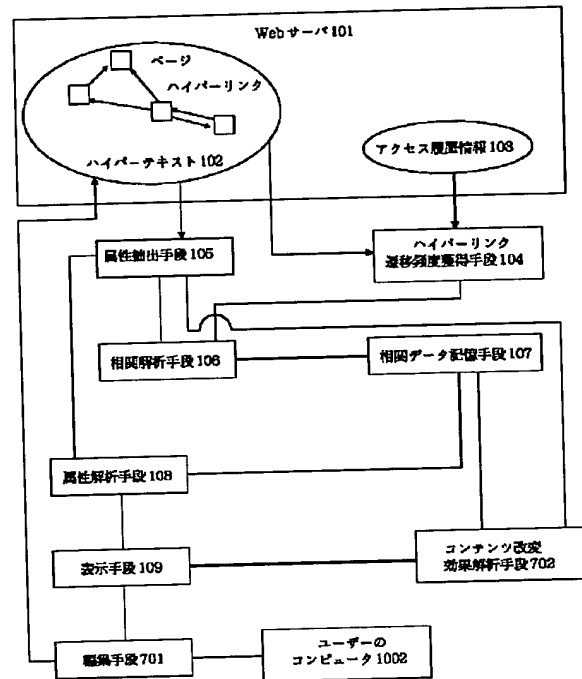
【図12】



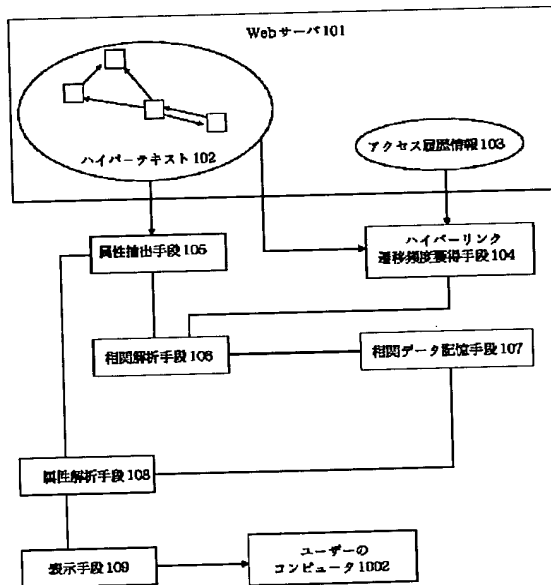
【図13】



【図14】



【図15】



【手続補正書】

【提出日】平成11年12月6日（1999.12.6）

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正内容】

【書類名】明細書

【発明の名称】ハイパーテキスト解析装置および方法

【特許請求の範囲】

【請求項1】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項2】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ群のうちハイパーリンクで繋がれたページ間について、該ハイパーリンク遷移頻度獲得手段を用いてハイパーリンク遷移頻度を計算し、該ハイパーリンク遷移頻度が小さなページ組を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、

どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項3】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ群のうちハイパーリンクで繋がれたページ間について、該ハイパーリンク遷移頻度獲得手段を用いてハイパーリンク遷移頻度を計算し、さらに、該属性抽出手段を用いて、該ページ間のコンテンツ類似度を計算し、該ハイパーリンク遷移頻度と該コンテンツ類似度の比率を基に、コンテンツが類似しているにもかかわらずハイパーリンク遷移頻度が小さなページ組を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項4】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出手段と、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項5】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、

該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出手段と、

該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ群の任意のページ間について、該アクセス類似度解析手段を用いてページ間アクセス類似度を計算し、該ページ間アクセス類似度が小さなページ組を獲得する解析対象ページ組獲得手段と、

該解析対象ページ組獲得手段で獲得されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項6】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、

該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出手段と、

該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ群の任意のページ間について、該アクセス類似度解析手段を用いてページ間アクセス類似度を計算し、さらに、該属性抽出手段を用いて、該ページ組のコンテンツ類似度を計算し、該ページ間アクセス類似度と該コンテンツ類似度の比率を基に、コンテンツが類似しているにもかかわらずページ間アクセス類似度が小さなページ組を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセ

ス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段とを有することを特徴とするハイパーテキスト解析装置。

【請求項7】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、

ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出手段と、

該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、

該属性解析手段で得られた結果を表示する表示手段と、この表示結果を参照しながら該指定されたページ組のコンテンツを改変する編集手段と、

該編集手段において改変されたコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を予測して改変による効果を計算するコンテンツ改変効果解析手段とを有し、

該表示手段において該コンテンツ改変効果解析手段で得られた結果を表示することを特徴とするハイパーテキスト解析装置。

【請求項8】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、

該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出手段と、

該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、

該相関解析手段で得られたデータを記憶する相関データ記憶手段と、

指定されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させることを想

定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段と、この表示結果を参照しながら該指定されたページ組のコンテンツを改変する編集手段と、該編集手段において改変されたコンテンツから該属性抽出手段を用いて属性を抽出し、これを該関連データ記憶手段において記憶されている関連データと比較することによって、該ページ組間のアクセス類似度を予測して改変による効果を計算するコンテンツ改変効果解析手段とを有し、該表示手段において該コンテンツ改変効果解析手段で得られた結果を表示することを特徴とするハイパーテキスト解析装置。

【請求項9】 請求項1及至請求項8に記載のハイパーテキスト解析装置において、関連解析手段が、ハイパーリンク遷移頻度と属性の相関、あるいは、アクセス類似度と属性の相関を計算する際に、有効ではないとおもわれる属性を選択する機能を有し、以下の処理においてこれらの属性を無視することを特徴とするハイパーテキスト解析装置。

【請求項10】 請求項1及至請求項9に記載のハイパーテキスト解析装置において、解析対象となるハイパーテキストシステムをネットワーク経由で指定する解析対象指定手段を有することを特徴とするハイパーテキスト解析装置。

【請求項11】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得ステップと、ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出ステップと、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析ステップと、該相関解析ステップで得られたデータを記憶する関連データ記憶ステップと、指定されたページ組のコンテンツから該属性抽出ステップにより属性を抽出し、これを該関連データ記憶ステップにより記憶されている関連データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析ステップと、該属性解析ステップで得られた結果を表示する表示ステップとを有することを特徴とするハイパーテキスト解析方法。

【請求項12】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析

ステップと、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出ステップと、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析ステップと、該相関解析ステップで得られたデータを記憶する関連データ記憶ステップと、指定されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出ステップにより属性を抽出し、これを該関連データ記憶ステップにより記憶されている関連データと比較することによって、該ページ組間のページ間アクセス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析ステップと、該属性解析ステップで得られた結果を表示する表示ステップとを有することを特徴とするハイパーテキスト解析方法。

【請求項13】 ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれたページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得ステップと、ハイパーリンクで繋がれたページ組のコンテンツから属性を抽出する属性抽出ステップと、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析ステップと、該相関解析ステップで得られたデータを記憶する関連データ記憶ステップと、指定されたページ組のコンテンツから該属性抽出ステップにより属性を抽出し、これを該関連データ記憶ステップにより記憶されている関連データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析ステップと、該属性解析ステップで得られた結果を表示する表示ステップとをコンピュータに実行させるために用いるプログラムを記録したハイパーテキスト解析用のコンピュータ読み取り可能な記録媒体。

【請求項14】 ハイパーテキストシステムを構成するページ組について、アクセス履歴情報を解析することによって、ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析ステップと、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから属性を抽出する属性抽出ステップと、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析ステップと、該相関解析ステップで得られたデータを記憶する関連データ記憶ステップと、

指定されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出ステップにより属性を抽出し、これを該関連データ記憶ステップにより記憶されている関連データと比較することによって、該ページ組間のページ間アクセス類似度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析ステップと、該属性解析ステップで得られた結果を表示する表示ステップとをコンピュータに実行させるために用いるプログラムを記録したハイパーテキスト解析用のコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ネットワーク上に構成されるハイパーテキストシステムにおいて、その構成に問題がある部分を改善する技術に関する。

【0002】

【従来の技術】ネットワーク上に構成されるハイパーテキストシステム（例えば、World Wide Web；以降Webと略す）では、ハイパーテキストを格納しているサーバにおいてユーザー（訪問者）のアクセス履歴を記録することができる。このアクセス履歴には、一般に、アクセスしてきたユーザーが使用しているコンピュータの識別子（インターネットを利用しているのであればIPアドレス）、アクセスしてきた時刻、アクセスしたページ（＝ファイル）のサーバ上での識別子（WebではURL）が含まれる。

【0003】アクセス履歴とハイパーリンク構造を解析して、ハイパーテキストシステム（例えば、Webサイト）の構成の優劣を判断するような知識を得る技術としては、[Perkowitz and Etzion, 98] (Perkowitz and Etzion, Adaptive Web Sites: Automatically Synthesizing Web Pages, in Proc. of AAAI-98) がある。

【0004】[Perkowitz and Etzion, 98] は、まず、サイト上の全てのページ組について、同一ユーザーからのアクセス共起頻度を計算し、予め定められた閾値を超える組を残し、その他を破棄する。さらに、残った組の中から実際にハイパーリンクで連結されている組を破棄する。次に、残った組をアークで連結されたグラフとみなし、各グラフを解析してクリーク（全てのノード間にアークが存在する完全グラフ）を抽出する。このクリークを構成するページ群は、互いに（強い）関連があるがハイパーリンクで連結されていないものなので、サイト上の劣っている部分を発見できたと解釈することができる。この手法では、多くのユーザーが同一セッションでアクセスしてくる傾向が強

いにもかかわらずそれらの間にハイパーリンクが存在しない（そのため、ユーザーはページ間遷移に多くの労力を求められるであろう）ページ群を発見できる。

【0005】

【発明が解決しようとする課題】従来技術では、構成の劣ったページ群を明示的に提示することはできるが、該ページ群の問題の原因を解析することはできなかった。そのため、ハイパーテキストシステム管理者は、ページ群の改善方法を試行錯誤によって探らなければならなかった。これは、ある部分を変更してみて、しばらくの間アクセス履歴を収集し、改善効果を測定するという操作を繰り返すもので、この間、ハイパーテキストシステムの構成は不安定となり、繰り返しアクセスしてくるユーザーを混乱させてしまうという問題が生じる。また、ハイパーテキストシステムにおけるページ群を改善するための手法は、そのハイパーテキストシステムの目的、規模、機能、レイアウト、話題、ユーザータイプ等に依存するので、対象ハイパーテキストシステムの特徴を把握しなければならないという課題もある。

【0006】本発明は上記事情を考慮してなされたもので、対象ハイパーテキストシステム上の任意のページ組について、ページコンテンツから抽出される様々な属性とページ間遷移頻度との相関を予め計算しておき、該ハイパーテキストシステム上で構成に問題がある部分についてどの属性をどのように変更すれば構成の改善になるかを提示することを目的とする。

【0007】

【課題を解決するための手段】本発明の第1の側面によれば、上述の目的を達成するために、ハイパーテキスト解析装置に、ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれた（例えば全ての）ページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから（一つ以上の）属性を抽出する属性抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する関連データ記憶手段と、指定された（例えば一つの）ページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該関連データ記憶手段において記憶されている関連データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させる（一般に、上げることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けている。

【0008】この構成においては、設計者がユーザーの遷移を期待してハイパーリンクを用いて関連付けたページ間においてどの属性を変化させれば初期の遷移頻度が得

られるかを簡易に提示することができる。ハイパーリンクシステム例えばWebシステムの管理者は、この提示内容に基づいてWebページの属性を変化させて所望のWebシステムを構築維持することができる。

【0009】また、本発明の第2の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれた（例えば全ての）ページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから（一つ以上の）属性を抽出する属性抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ群のうちハイパーリンクで繋がれた（例えば全ての）ページ間について、該ハイパーリンク遷移頻度獲得手段を用いてハイパーリンク遷移頻度を計算し、該ハイパーリンク遷移頻度が小さなページ組（例えば一つ以上）を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を变化させる（一般に、上げることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けるようにしている。

【0010】この構成においても、ハイパーリンクを用いて関連付けたページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。さらに、ハイパーリンクでリンクされたページでありながら遷移頻度の少ないページを修正対象として自動的に選別することができる。

【0011】また、本発明の第3の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造とを解析して、ハイパーリンクで繋がれた（例えば全ての）ページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから（一つ以上の）属性を抽出する属性抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ群のうちハイパーリンクで繋がれた（例えば全ての）ページ間について、該ハイパーリンク遷移頻度獲得手段を用いてハイパーリンク遷移頻度を計算し、さらに、該属性抽出手段を用いて、該ページ間のコンテ

ツ類似度を計算し、該ハイパーリンク遷移頻度と該コンテンツ類似度との比率を基に、コンテンツが類似しているにもかかわらずハイパーリンク遷移頻度が小さなページ組（例えば一つ以上）を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を变化させる（一般に、上げることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けるようにしている。

【0012】この構成においても、ハイパーリンクを用いて関連付けたページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。さらに、ハイパーリンクでリンクされたページであり、かつ関連したコンテンツを有しながら遷移頻度の少ないページを修正対象として自動的に選別することができる。

【0013】また、本発明の第4の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムを構成する（例えば全ての）ページ組について、アクセス履歴情報を解析することによって、（多くの）ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから（一つ以上の）属性を抽出する属性抽出手段と、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定された（例えば一つの）ページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を变化させる（一般に、大きくすることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けるようにしている。

【0014】この構成においては、リンクにより直接繋がっていないページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。

【0015】また、本発明の第5の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムを

構成する（例えば全ての）ページ組について、アクセス履歴情報を解析することによって、（多くの）ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから（一つ以上の）属性を抽出する属性抽出手段と、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ群の任意のページ間について、該アクセス類似度解析手段を用いてページ間アクセス類似度を計算し、該ページ間アクセス類似度が小さなページ組（一つ以上）を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させる（一般に、大きくすることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けるようにしている。

【0016】この構成においても、リンクにより直接繋がっていないページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。さらに、ユーザによるアクセスが共起することの少ないページ間を修正対象として自動的に選別することができる。

【0017】また、本発明の第6の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムを構成する（例えば全ての）ページ組について、アクセス履歴情報を解析することによって、（多くの）ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから（一つ以上の）属性を抽出する属性抽出手段と、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定されたページ群の任意のページ間について、該アクセス類似度解析手段を用いてページ間アクセス類似度を計算し、さらに、該属性抽出手段を用いて、該ページ間のコンテンツ類似度を計算し、該ページ間アクセス類似度と該コンテンツ類似度の比率を基に、コンテンツが類似しているにもかかわらずページ間アクセス類似度が小さなページ組（一つ以上）を獲得する解析対象ページ組獲得手段と、該解析対象ページ組獲得手段で獲得されたページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該

属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させる（一般に、大きくすることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段とを設けるようにしている。

【0018】この構成においても、リンクにより直接繋がっていないページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。さらに、コンテンツに関連しているながらユーザによるアクセスが共起することの少ないページ間を修正対象として自動的に選別することができる。

【0019】また、本発明の第1～第6の側面において、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるハイパーリンクの位置を抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるハイパーリンクの数を抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるハイパーリンクの種類を抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるハイパーリンクのサイズを抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるハイパーリンクを表わす文字の種類を抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページのコンテンツにおけるページのサイズを抽出するようにしてもよい。

【0020】また、本発明の第1、第2、第4または第5の側面において、属性抽出手段が、属性の一つとして、少なくとも、ページ組間のコンテンツ類似度を抽出するようにしてもよい。

【0021】また、本発明の第1～第6の側面において、属性抽出手段が、属性の一つとして、少なくとも、ページの位置を抽出するようにしてもよい。また、属性抽出手段が、属性の一つとして、少なくとも、ページの更新日時を抽出するようにしてもよい。

【0022】また、本発明の第4～第6の側面において、属性抽出手段が、属性の一つとして、少なくとも、ページ間のハイパーリンク遷移数を抽出するようにしてもよい。

【0023】また、本発明の第1～第3の側面において、ハイパーリンク遷移頻度解析手段が、ハイパーテキストを網羅的にアクセスして自動的に情報を収集する探索ロボットを識別して、当該探索ロボットからのアクセ

ス情報を排除して解析を行うようにしてもよい。

【0024】また、本発明の第4～第6の側面において、アクセス類似度解析手段が、ハイパーテキストを網羅的にアクセスして自動的に情報を収集する探索ロボットを識別して、当該探索ロボットからのアクセス情報を排除して解析を行うようにしてもよい。

【0025】また、本発明の第1～第3の側面において、ハイパーリンク遷移頻度解析手段が、ブロッキシャーバーを識別して、当該ブロッキシャーバからのアクセス情報を排除して解析を行うようにしてもよい。

【0026】また、本発明の第4～第6の側面において、アクセス類似度解析手段が、ブロッキシャーバーを識別して、当該ブロッキシャーバからのアクセス情報を排除して解析を行うようにしてもよい。

【0027】また、本発明の第1～第3の側面において、ハイパーリンク遷移頻度解析手段が、被リンク数が大きなページをリンク先とするハイパーリンク遷移を排除して解析を行うようにしてもよい。

【0028】また、本発明の第7の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムへのアクセス履歴情報とハイパーリンク構造を解析して、ハイパーリンクで繋がれた（例えば全ての）ページ間のハイパーリンク遷移頻度を計算するハイパーリンク遷移頻度獲得手段と、ハイパーリンクで繋がれたページ組のコンテンツから（一つ以上の）属性を抽出する属性抽出手段と、該ハイパーリンク遷移頻度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定された（例えば一つの）ページ組のコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を変化させる（一般に、上げることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段と、この表示結果を参照しながら該指定されたページ組のコンテンツを改変する編集手段と、該編集手段において改変されたコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のハイパーリンク遷移頻度を予測して改変による効果を計算するコンテンツ改変効果解析手段とを設け、該表示手段において該コンテンツ改変効果解析手段で得られた結果を表示するようにしている。

【0029】この構成においては、ハイパーリンクを用いて関連付けたページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、また属性を調整した後の予想結果に基づいて確実に属性調整を行え、所望のハイパーリンクシステム例え

ばWebシステムを構築維持することができる。

【0030】また、本発明の第8の側面によれば、ハイパーテキスト解析装置に、ハイパーテキストシステムを構成する（例えば全ての）ページ組について、アクセス履歴情報を解析することによって、（多くの）ユーザーが双方をアクセスした度合いを表わすページ間アクセス類似度を計算するアクセス類似度解析手段と、該ページ組のコンテンツや該ページ組を包含するハイパーテキストから（一つ以上の）属性を抽出する属性抽出手段と、該ページ間アクセス類似度と該属性との間の相関を計算する相関解析手段と、該相関解析手段で得られたデータを記憶する相関データ記憶手段と、指定された（例えば一つの）ページ組のコンテンツや該ページ組を包含するハイパーテキスト構造から該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のページ間アクセス類似度を変化させる（一般に、大きくすることを目的とする）ことを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報を求める属性解析手段と、該属性解析手段で得られた結果を表示する表示手段と、この表示結果を参照しながら該指定されたページ組のコンテンツを改変する編集手段と、該編集手段において改変されたコンテンツから該属性抽出手段を用いて属性を抽出し、これを該相関データ記憶手段において記憶されている相関データと比較することによって、該ページ組間のアクセス類似度を予測して改変による効果を計算するコンテンツ改変効果解析手段とを設け、該表示手段において該コンテンツ改変効果解析手段で得られた結果を表示するようにしている。

【0031】この構成においては、リンクにより直接繋がっていないページ間においてどの属性を変化させれば初期の遷移頻度が得られるかを簡易に提示することができ、また属性を調整した後の予想結果に基づいて確実に属性調整を行え、所望のハイパーリンクシステム例えばWebシステムを構築維持することができる。

【0032】また、本発明の第1～第6の側面において、相関解析手段が、ハイパーリンク遷移頻度と属性との間の相関（あるいは、アクセス類似度と属性との間の相関）を計算する際に、有効ではないとおもわれる属性を選択する機能を有し、後続する処理においてこれらの属性を無視するようにしてもよい。

【0033】また、本発明の第1～第6の側面において、解析対象となるハイパーテキストシステムを（例えばネットワーク経由で）指定する解析対象指定手段を設けるようにしてもよい。

【0034】また、本発明の第1～第6の側面において、複数のユーザーが、解析対象となるハイパーテキストシステムを（例えばネットワーク経由で）指定する解析対象指定手段を設けるようにしてもよい。

【0035】また、本発明の第1～第6の側面において、解析対象となるハイパーテキストシステムのコンテンツを（例えばネットワーク経由で）受け渡すコンテンツ伝送手段を設けるようにしてもよい。

【0036】また、本発明の第1～第6の側面において、複数のユーザーが、解析対象となるハイパーテキストシステムのコンテンツを（例えばネットワーク経由で）受け渡すコンテンツ伝送手段を有するようにしてもよい。

【0037】また、本発明の第1～第6の側面において、解析対象となるハイパーテキストシステムのアクセス履歴情報を（例えばネットワーク経由で）受け渡すアクセス履歴情報伝送手段を設けるようにしてもよい。

【0038】また、本発明の第1～第6の側面において、複数のユーザーが、解析対象となるハイパーテキストシステムのアクセス履歴情報を（例えばネットワーク経由で）受け渡すアクセス履歴情報伝送手段を有するようにしてもよい。

【0039】また、本発明の第1～第6の側面において、属性抽出手段が抽出する属性の組を（例えばネットワーク経由で）指定する属性指定手段を設けるようにしてもよい。

【0040】また、本発明の第1～第6の側面において、複数のユーザーが、属性抽出手段が抽出する属性の組を（ネットワーク経由で）指定する属性指定手段を有するようにしてもよい。

【0041】また、本発明の第7の側面において、編集手段がネットワーク経由で操作できるようにしてもよい。また、編集手段がネットワーク経由で複数のユーザーによって操作できるようにしてもよい。

【0042】また、本発明の第1～第6の側面において、表示手段が表示内容をネットワーク経由で伝送できるようにしてもよい。また、表示手段が表示内容をネットワーク経由で複数のユーザーに対して伝送できるようにしてもよい。

【0043】なお、本発明は方法の態様で実現することもできる。この場合、ハイパーリンク解析方法は、ハイパーリンク解析装置を構成する各手段に対応する手順で構成される。また、本発明の構成のうちの少なくともその一部をコンピュータソフトウェアとして実現することも可能である。また本発明を当該コンピュータソフトウェアを記録した記録媒体により実現することも可能である。

【0044】

【発明の実施の形態】ネットワーク上に構成されるハイパーテキストシステムのなかで代表的なものはWebである。以下、HTML（ハイパーテキストマークアップランゲージ）で記述されたWebを例として、本発明の実施例を説明する。

【0045】〔実施例1〕図1は、本発明の実施例1を

示す構成図である。同図において、Webサーバ101はネットワーク上で情報を発信する手段であり、Webサーバ101にはユーザーに提供したい情報が、ページ（＝ファイル）とハイパーリンクからなるハイパーテキスト102として貯えられている。ユーザーはWebサーバ101上のハイパーテキスト102にアクセスすることで情報を入手できる。Webサーバ101は、ユーザーからのアクセスがある毎に、ユーザーのコンピュータを識別するためのコンピュータ識別子（IPアドレス）とアクセス時刻とユーザーのアクセスしたページのあるアドレス（URL）をアクセス履歴情報103として記録する手段を持つ。

【0046】ハイパーリンク遷移頻度獲得手段104は、ハイパーテキスト102の各ページ（起点ページと呼ぶ）について、そこからのリンク先ページ（Webサーバ101上にあるものだけを対象とし、Webサーバ101の外にあるものは無視する）を獲得して、それぞれのページ組について、アクセス履歴情報103を参照して、同一ユーザーから同一セッションで起点ページからリンク先ページの順にアクセスがあれば一件として順次カウントしていく（ある定められた時間枠で、各IPアドレス毎にアクセス先ページ（URL）を時間順に並べることによって容易に実現できる）。得られた積算カウントがページ間（つまり、ハイパーリンクの）遷移頻度である。ここで、本発明の変形例として、積算カウントを起点ページへのアクセス数で割った値をハイパーリンク遷移頻度として扱ってもよい。また別の変形例として、ハイパーリンク遷移頻度獲得手段104が、情報収集のための探索ロボットからのアクセスを無視するようにしてもよい。これは、例えば、ある慣習に従って探索ロボットがアクセスする特殊なページ（Webでは、ルート直下に置かれるrobots.txtという名のファイル）へのアクセスの有無や、短期間に多数のノードを網羅的にアクセスするという探索ロボットに特徴的な振る舞いの有無や、既知の探索ロボットのコンピュータ識別子であるか否かによって、探索ロボットを識別して、該探索ロボットのコンピュータ識別子に関わる情報を排除することによって実現できる。こうすることによって、ハイパーリンク遷移頻度がより正確に実ユーザー（＝人間）の振る舞いを反映したものとなり、本ハイパーテキスト解析装置の精度が向上する。さらに別の変形例として、ハイパーリンク遷移頻度獲得手段104が、プロクシーサーバからのアクセスを無視するようにしてもよい。これは、アクセス履歴情報103におけるコンピュータ識別子のうち、プロクシーサーバに対応するのは（一般に）複数のユーザーからのアクセスを集約したものなので、そのハイパーリンク遷移頻度を求めても個別のユーザーの振る舞いを正確に反映したものにはならず解析精度が下がってしまうという問題を解決するために行うものである。プロクシーサーバであるか否か

は、ある期間に、予め定めておいた閾値（一個人が通常アクセスすることがないようなアクセス頻度を閾値とする）を越える数のページにアクセスしたかどうかや、既知のプロキシサーバのコンピュータ識別子であるか否かによって識別できる。さらに別の変形例として、ハイパーリンク遷移頻度獲得手段104が、ハイパーテキスト102のハイパーリンク構造を解析して各ページの被リンク数を求め、その値が予め定められた閾値を越えるようなページへのハイパーリンク遷移を無視するようにしてもよい。リンク数または被リンク数が大きなページはサイトナビゲート用のインデックスページであることが多い。これらのページはサイトのコンテンツ提供者にとってユーザーにアクセスしてもらいたいコンテンツを包含するページというよりも、サイトの性能を維持するための機能的役割を果たすためのものである傾向が強く、これらをリンク先とするハイパーリンク遷移頻度を求めても、コンテンツを探索するユーザーの振る舞いを直接的に定量化することにはつながらない。そこで、これらのページをリンク先とするハイパーリンク遷移を無視することによって、解析精度が上がり、さらに計算量も減少して処理効率が良くなる。

【0047】属性抽出手段105は、ハイパーテキスト102の各ページから（一つ以上の）属性を抽出する。属性としては、例えば、以下の9項目の中から任意のN個の組み合わせ（Nは1以上9以下）を抽出する。もちろん他の属性を用いてもよい。

【0048】【属性1】ページのコンテンツにおける、それぞれのリンク先ページに対応するハイパーリンクの位置を抽出する。これは、ハイパーリンクが埋め込まれている位置がページのコンテンツの先頭から何文字目か（文字カウントでわかる）、あるいは、何単語目か（既存の形態素解析技術を用いることによって実現できる）、あるいは、何パラグラフ目か（タグ情報を用いることによって実現できる）という情報を得ることによって実現できる。また、本発明の変形例として、ハイパーリンクの位置を、ページがブラウザでレンダリングされ、表示される際のウインドウ枠内の位置として定めてもよい。これは、標準となるウインドウ枠サイズを予め定めておき、ハイパーリンクがウインドウ枠内に表示されるかどうか、あるいは、ウインドウ枠を例えば縦に均等3分割して（上部、中部、下部、見えない部分）と領域を定め、ハイパーリンクがどの領域に属するかを抽出する。また、別の変形例として、ハイパーリンクの位置を、ページコンテンツの全てのハイパーリンクの中で何番目に現れるかという順序で定めてもよい。

【0049】【属性2】ページのコンテンツにおけるハイパーリンクの数を抽出する。また、本発明の変形例として、特定のページ（＝注目しているリンク先ページ）を指すハイパーリンクの数としてもよい。これは、ハイパーテキスト102の全てのページからハイパーリンク

を抽出してそのリンク先URLを調べることによって容易にカウントできる。

【0050】【属性3】ページのコンテンツにおけるハイパーリンクがテキストで示されているか、あるいは、画像で示されているかを抽出する。これは、ページのコンテンツからハイパーリンクを表わすタグを取り出し（`...`）、この中に画像ファイルが埋め込まれている（``）か否かを調べることによって実現できる。

【0051】【属性4】ページのコンテンツにおけるハイパーリンクのサイズを抽出する。サイズの尺度としては、例えば、テキストであれば文字数や単語数（既存の形態素解析技術を用いることによって実現できる）、画像であれば表示領域の面積（`IMG`タグ中の`width`と`height`の値を掛け合わせて求める）を用いる。

【0052】【属性5】ページのコンテンツにおけるハイパーリンクテキストの文字種を抽出する。文字種としては、例えば、フォントサイズ、あるいは、フォント色、イタリックやボールド等の文字飾りの有無を抽出する。これらは、タグによって規定されているので容易に抽出できる。

【0053】【属性6】ページのサイズを抽出する。サイズの尺度としては、例えば、コンテンツの文字数や単語数（既存の形態素解析技術を用いることによって実現できる）、あるいはページファイルのサイズ（Byte数）を用いる。

【0054】【属性7】起点ページとリンク先ページのコンテンツ類似度を抽出する。コンテンツ類似度の計算法は、既存技術を用いる。例えばベクトル空間モデル法 [Salton and Allan, 1994] (Salton and Allan, Text Retrieval Using the Vector Processing Model, in Proc. of SDAIR-94.) では、各ページについて、単語を項としその出現頻度を項の値とするようなベクトルを生成して、ベクトル間の内積値をコンテンツ類似度として用いている。ここで、本発明の変形として、各起点ページのハイパーリンクが埋め込まれている位置の近傍テキスト（例えば、同一パラグラフとか前後N単語を近傍とする）に出現する単語だけを対象とする（リンク先ページは、ページ全体に出現する単語を対象とする、あるいは、ハイパーリンクにリンク先のページ中の位置まで記述してあれば、その近傍テキストを対象とする）ようにしてもよい。

【0055】【属性8】ページ間の論理的な位置差を抽出する。例えば、Webサーバ101上でページのファイルが置かれているディレクトリの位置を論理的な位置として、ルートからの深さを抽出する。次に、起点ページ

とリンク先ページの深さの差を計算してその値を論理的
位置差分とする（ルートからの深さは、URLを解析す
ればわかる。例えば、起点ページがPage1.html
という名前のファイルで、www.○○.co.jp/
Dir1/Page1.htmlというURLで表わ
されていて、リンク先ページがPage2.htmlと
いう名前のファイルで、www.○○.co.jp/
Dir1/Dir2/Page2.htmlというURL
で表わされている場合、論理的位置差分は1である）。

【0056】[属性9] ページのファイルの更新日時を
調べ（更新日時は、Webサーバ101においてファ
イルプロパティを調べれば容易にわかる）、本発明のハイ
パーテキスト解析装置使用時と更新日時の差分時間を属
性として抽出する。

【0057】図2は、各ページについて、ハイパーリン
ク遷移頻度獲得104で得られたハイパーリンク遷移頻
度と、属性抽出手段105で得られた属性を示した表の
例の一部である。この例では、ハイパーリンク遷移頻度
として、積算カウントを起点ページへのアクセス数で割
った値を用い、属性の組合わせとして、上記の1（先頭
からの単語数カウントを使用）、4（文字数を使用）、
6（起点ページの単語数を使用）の3つを用いている。

【0058】図1の相関解析手段106では、ハイパー
リンク遷移頻度獲得手段104で得られたハイパーリン
ク遷移頻度と、属性抽出手段105で得られた属性の相
関を計算する。相関の計算法としては、既存の統計手法
を用いることができるが、ここでは、重回帰分析を用い
た場合の実施例を説明する。本発明に重回帰分析を適用
した場合、目的変量はハイパーリンク遷移頻度（ y と
おく）で、説明変量は各属性（ x_1, x_2, x_3, \dots
 x_n とおく）となる（図2の例では、目的変量は遷移頻
度（ y ）で、説明変量はリンク位置（ x_1 ）とリンクサ
イズ（ x_2 ）とページサイズ（ x_3 ）の3つとなる）。
重回帰分析では、 $y = a_1 x_1 + a_2 x_2 + a_3 x_3 + \dots + a_n x_n + a_0$ （ a は回帰係数）という重
回帰式を作り、左辺と右辺の y と x にデータを代入し
て、その差の二乗を表わす式を得る。一般にデータ系列
は複数（多数）あるので、それぞれについて左辺と右辺
の差の二乗を表わす式を得て、全ての式の和（ Q と
おく）を得る。次に、 Q を最小にするような回帰係数を求
めるという手続きを取る（ Q をそれぞれの回帰係数で偏
微分してその値を0とおいた等式を作り、その連立方程
式を解けばよい）。こうして得られた重回帰式は、図1
の相関データ記憶手段107において記憶される。こ
こで、本発明の変形例として、説明変量の候補の中から、
目的変量（＝ハイパーリンク遷移頻度）の予測に有効で
はないと思われる属性を選択し、以下の処理においてこ
れらの属性を無視することによって、解析精度を上げ、
さらに処理効率を上げるようにしてもよい。重回帰分析
において、有効な説明変量の組合わせを選択する手法は

既存であり、例えば、forward selection methodやstepwise method
が提案されている（Carl J. Huberty, Applied Discriminant Analysis, John Wiley & Sons, Inc, 1994）ので、本発明ではこれらのうちいずれか
を用いれば容易に実現できる。

【0059】次に、あるページ組が解析対象として与え
られると、本ハイパーリンク解析装置は、属性抽出手段
105を用いて、前述の方法により、属性を抽出する
（上で、説明変量として採用した属性だけを抽出すれば
よい。図2の例では、上記の1、4、6の3つの属性を
抽出する）。このようにして得られた属性は、属性解析
手段108において、相関データ記憶手段107におい
て記憶されている重回帰式と比較される。その方法は、
まず、ある一つの属性に注目して（例えば、図2の例
で、上記の1の属性に注目したとする）、他の属性の値
を重回帰式に代入する（上記の4と6の属性の値をそれ
ぞれ x_4 と x_6 に代入する）。すると、回帰式は、一変
数の一次式で表わされ（ $y = a_1 x_1 + c$ （ c は定数）
という形になる）、与えられたページ組のハイパーリン
ク遷移頻度と注目した属性の（予測）相関式が得られる
ので、（他の属性は変化させないという前提の基におい
て）該属性をどういう値に変更すれば、ハイパーリンク
遷移頻度がどのように変わるかが予測される。尚、ここ
で、注目している属性の値（属性抽出手段105で得ら
れたもの）を代入すると、現状のハイパーリンク遷移頻
度予測値が得られる。本発明では、上記の注目している
属性というのは、特に指定されるものではなく、全ての
属性を順に注目しながら、上記の処理を繰り返す（つま
り、属性の数だけ相関式が得られる）。

【0060】このようにして得られた相関式は表示手段
109において表示される。図3は、表示の一例を示す
ものである。ここでは、図2の例の結果表示の一部（上
記の1の属性に注目したものを示している）を示してい
る。同図では、与えられたページ組のハイパーリンク遷
移頻度を注目した属性（ハイパーリンクの位置）の（一
次）関数として表わしている。ここで、 x 軸の値 p は、
現状の属性値であり、 $q (= a_1 p + c)$ は、現状のハイ
パーリンク遷移頻度予測値となる。同図を参照する
と、属性値をどれだけ可変させると（ここでは、ハイパ
ーリンクの位置を何単語目に埋め込むかということにな
る）、どれだけハイパーリンク遷移頻度を可変させられ
るかが定量的に予測できる（一般に、ハイパーリンク遷
移頻度を高くすることを考えるが、ここでは、ハイパー
リンク位置を前にするほど、ハイパーリンク遷移頻度が
高くなることが予測でき、その効果は回帰係数 a_1 の値
で左右される）。

【0061】[実施例2] つぎに本発明の実施例2につ
いて説明する。

【0062】図4は、本発明の実施例2を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、解析対象ページ組獲得手段401により、属性解析手段108で解析するページ組を自動的に取捨選択する。この処理の流れは、先ず、指定されたページ集合（例えば、ある特定のサーバ上の全てのページ、とか、ある特定のサーバ上のある特定のディレクトリ以下に存在する全てのページといった指定）のうちハイパーリンクで直接繋がれたページ組について、ハイパーリンク遷移頻度獲得手段104を用いてハイパーリンク遷移頻度を得る。次に、この情報は解析対象ページ組獲得手段401に渡され、解析対象ページ組獲得手段401において、ハイパーリンク遷移頻度が予め定められた閾値より小さなページ組を獲得して、これらを属性解析手段108に渡す。属性解析手段108では、上記の実施例の方法で、渡されたページ組の解析を（順次）行う。こうすることによって、本ハイパーテキスト解析装置のユーザーが、予め、（ハイパーリンク遷移頻度）に問題がありそうなページ組を探して指定しなければならないという手間を省略できる。

【0063】ここで、本発明の変形例を示す。先ず上記の処理と同様に、指定されたページ集合のうちハイパーリンクで直接繋がれたページ組について、ハイパーリンク遷移頻度獲得手段104を用いてハイパーリンク遷移頻度を得る。さらに、本変形例では、属性抽出手段105において、前述の方法により該ページ組間のコンテンツ類似度を得る。こうして得られた該ページ組間のハイパーリンク遷移頻度とコンテンツ類似度の情報は解析対象ページ組獲得手段401に渡され、解析対象ページ組獲得手段401において、ハイパーリンク遷移頻度をコンテンツ類似度で割った値が予め定められた閾値より小さなページ組を獲得して、これらを属性解析手段108に渡す。ハイパーリンク遷移頻度をコンテンツ類似度で割った値が小さなページ組は、内容が類似しているにも拘わらず、ユーザーがリンクを遷移してくれる傾向が小さいものであり、改善の余地があるといえる（改善のための指針は上記と同様に属性解析手段108において得られ、表示手段109において表示される）。

【0064】〔実施例3〕つぎに本発明の実施例3について説明する。

【0065】図5は、本発明の実施例3を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。同図において、アクセス類似度解析手段501は、ハイパーテキスト102の（全ての）ページ組について、そのページ間のアクセス類似度を求める。ここで、アクセス類似度とは、与えられた2つのページについて、いかに多くのユーザーが双方をアクセスしたかという度合いを表わす指標であり、例えば、ベクトル空間モデル法を使って、各ページについて、アクセスのあったIPアドレスを項としその出現頻

度を項の値とするようなベクトルを生成して、ベクトル間の内積値をアクセス類似度として用いる。このように前述のハイパーリンク遷移頻度の代わりに、アクセス類似度を用いることによって、（2つのページ間のハイパーリンクの有無を考慮することなく）任意のページ組を対象として、そのページ構成の改善策を提示できるようになる。また本発明の変形例としてアクセス類似度解析手段501が、上記実施例と同様の方法で、情報収集のための探索ロボットからのアクセスを無視するようにしてもよい。さらに別の変形例として、アクセス類似度解析手段501が、上記実施例と同様の方法で、プロクシーサーバからのアクセスを無視するようにしてもよい。

【0066】属性抽出手段105は、上記実施例と同様の処理を行うが、ここでは、さらに、ページ間のハイパーリンク遷移数を計算してこれを属性の1つとしてもよい。具体的には、ページ組（ページA、ページB）が与えられたときに、ハイパーリンクの向きを考慮しながらページAからページBへの最短経路の遷移数を計算し、次にページBからページAへの最短経路の遷移数を計算し、両者を比べて小さい方をページ組（ページA、ページB）のハイパーリンク遷移数とする。ここで、ページAからページBへの最短経路は、ページAからページBに到達するまで幅優先で深さ（＝遷移数）をカウントしながらハイパーリンクを探索すればよい。この探索は、一般には、辿ることができるハイパーリンクが無くなるまで続けられるが、処理効率を上げるために、深さ（＝遷移数）が予め決められた閾値に達したところで停止するように設定してもよい。この場合、ページ間の最短経路の遷移数は該閾値とする。

【0067】相関解析手段106では、アクセス類似度解析手段501で得られたアクセス類似度と、属性抽出手段105で得られた属性の相関を上記実施例と同様に重回帰分析によって計算する。得られた重回帰式は、相関データ記憶手段107において記憶される。ここで、本発明の変形例として、上記実施例と同様の方法により、説明変数の候補の中から、目的変数（＝アクセス類似度）の予測に有効ではないとおもわれる属性を選択し、以下の処理においてこれらの属性を無視することによって、解析精度を上げ、さらに処理効率を上げるようにしてもよい。

【0068】次に、あるページ組が解析対象として与えられると、本ハイパーリンク解析装置は、属性抽出手段105を用いて、上記実施例と同様の方法により、属性を抽出する。得られた属性は、属性解析手段108において、相関データ記憶手段107において記憶されている重回帰式と比較され、上記実施例と同様に、与えられたページ組のアクセス類似度と注目した属性の（予測）相関式が得られるので、（他の属性は変化させないという前提の基において）該属性をどういう値に変更すれば、ハイパーリンク遷移頻度がどのように変わるかが予

測される。尚、ここで、注目している属性の値（属性抽出手段105で得られたもの）を代入すると、現状のアクセス類似度予測値が得られる。本発明では、上記の注目している属性というのは、特に指定されるものではなく、全ての属性を順に注目しながら、処理を繰り返す（つまり、属性の数だけ相関式が得られる）。このようにして得られた相関式は、上記実施例と同様に、表示手段109において表示される。

【0069】【実施例4】つぎに本発明の実施例4について説明する。

【0070】図6は、本発明の実施例4を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、解析対象ページ組獲得手段601により、属性解析手段108で解析するページ組を自動的に取捨選択する。この処理の流れは、先ず、指定されたページ集合（例えば、ある特定のサーバ上の全てのページ、とか、ある特定のサーバ上のある特定のディレクトリ以下に存在する全てのページといった指定）から任意の2つを取り出したページ組について、アクセス類似度解析手段501を用いてアクセス類似度を得る。次に、この情報は解析対象ページ組獲得手段601に渡され、解析対象ページ組獲得手段601において、アクセス類似度が予め定められた閾値より小さなページ組を獲得して、これらを属性解析手段108に渡す。属性解析手段108では、上記の実施例の方法で、渡されたページ組の解析を（順次）行う。こうすることによって、本ハイパーテキスト解析装置のユーザーが、予め、（アクセス類似度）に問題がありそうなページ組を探して指定しなければならないという手間を省略できる。

【0071】ここで、本発明の変形例を示す。先ず上記の処理と同様に、指定されたページ集合から任意の2つを取り出したページ組について、アクセス類似度解析手段501を用いてアクセス類似度を得る。さらに、本変形例では、属性抽出手段105において、前述の方法により該ページ組間のコンテンツ類似度を得る。こうして得られた該ページ組間のアクセス類似度とコンテンツ類似度の情報は解析対象ページ組獲得手段601に渡され、解析対象ページ組獲得手段601において、アクセス類似度をコンテンツ類似度で割った値が予め定められた閾値より小さなページ組を獲得して、これらを属性解析手段108に渡す。アクセス類似度をコンテンツ類似度で割った値が小さなページ組は、内容が類似しているにも拘わらず、ユーザーがリンクを遷移してくれる傾向が小さいものであり、改善の余地があるといえる（改善のための指針は上記と同様に属性解析手段108において得られ、表示手段109において表示される）。

【0072】【実施例5】つぎに本発明の実施例5について説明する。

【0073】図7は、本発明の実施例5を示す構成図で

ある。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。尚、本発明では、同図のハイパーリンク遷移頻度獲得手段104をアクセス類似度解析手段501と置き換えることもできる。この場合、以下の説明中のハイパーリンク遷移頻度はアクセス類似度に置き換えられる。

【0074】本実施例は、編集手段701において、本ハイパーテキスト解析装置の操作者が、表示手段109で表示されている情報（指定されたページ組間のハイパーリンク遷移頻度を変化させることを想定した場合に、どの属性をどれだけ変化させれば、どれだけ効果があるかという情報）を参照しながら指定されたページ組のコンテンツを改変する（＝1つ以上の属性を変化させる）ことができる。図8は、表示手段109における表示の一例である。この例では、P6（起点ページ）とP7（リンク先ページ）が指定されたページ組であり、目的変数 y に対応するハイパーリンク遷移頻度には、積算カウントを起点ページへのアクセス数で割った値を用いている（アクセス類似度をに置き換えた場合は、上記実施例と同様に、IPアドレスの出現頻度に基づくベクトルの内積値を用いる）。また、属性の組合せとして、上記の1（ハイパーリンクの位置：先頭からの単語数カウントを使用）、4（ハイパーリンクサイズ：文字数を使用）、6（ページサイズ：起点ページの単語数を使用）の3つを用いており、これらが順に説明変数 x_1 、 x_2 、 x_3 にそれぞれ対応する。同図の表では、ページ組{P6, P7}から（属性抽出手段105において）抽出された3つの属性値がそれぞれ、 $x_1 = 66$ 、 $x_2 = 7$ 、 $x_3 = 325$ であり、そのとき予測されるハイパーリンク遷移頻度が $y = 0.17$ である（属性解析手段108において求められる）ことがわかる。該操作者は、この $y = 0.17$ というハイパーリンク遷移頻度予測値を増加させたいければ、同図下の3つのグラフを見て、どの属性をどれだけ変化させれば、どれだけ効果があるかということを知ることができる。本例では、 x_1 を小さくする（つまり、P6からP7へのハイパーリンクの位置を前にする）、あるいは、 x_2 を大きくする（つまりP6からP7へのハイパーリンクの文字数を増やす）、あるいは、 x_3 を小さくする（つまりP6の文字数を減らす）ことによってハイパーリンク遷移頻度予測値が増加することがわかる。このとき、各グラフの傾きを見ることによって属性値の可変量がどの程度の効果を持つかも容易に知ることができる（尚、これら3つのグラフはそれぞれに対応する属性だけを変化させたときのハイパーリンク遷移頻度予測値を示すものである）。該操作者は、これらの情報を参考にしながら、編集手段701において、ページ組{P6, P7}のコンテンツを改変できる。このようにして改変されたページ組はハイパーテキスト102に置かれる。

【0075】次に、属性抽出手段105において、上記

実施例と同様の方法で、該ページ組の属性が抽出され、コンテンツ改変効果解析手段702に渡される。コンテンツ改変効果解析手段702では、上記実施例と同様の方法で、該ページ組の属性と関連データ記憶手段において記憶されている関連データと比較することによって、該ページ組間のハイパーリンク遷移頻度を予測して改変による効果を計算し、その計算結果を表示手段109に渡す。

【0076】図9は、表示手段109における表示の一例である。同図の表は、ページ組{P6, P7}のコンテンツ改変後のハイパーリンク遷移頻度予測値を示すものである。本例では、該操作者は、編集手段701において x_1 と x_3 を変化させており、改変後のハイパーリンク遷移頻度予測値は0.17から0.31に増加している。本発明では、該操作者はこの表示例を参照しながら、編集手段701を用いて、さらに、コンテンツの改変を繰り返すこともできる。このとき、上記実施例と同様に、同図下の3つのグラフを見て、どの属性をどれだけ変化させれば、どれだけ効果があるかということを知ることができる。尚、これら3つのグラフの形状は、図8のコンテンツ改変前のものと比較して、回帰係数 a_1 、 a_2 、 a_3 は同一で、定数項だけが変化したものとなる。

【0077】[実施例6] つぎに本発明の実施例6について説明する。

【0078】図10は、本発明の実施例6を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、解析対象指定手段1001により、本ハイパーテキスト解析装置のユーザーが、解析対象となるWebサーバを指定できるようにする。ユーザーのコンピュータ1002と解析対象指定手段1001はネットワーク（例えばインターネットや電話回線）で接続されている。ユーザーのコンピュータ1002では、解析対象とするWebサーバのURL（サーバを識別する符号であれば何でもよい）が入力され、このURLはネットワークを通じて解析対象指定手段1001に渡される。解析対象指定手段1001は、該URLを属性抽出手段105とハイパーリンク遷移頻度獲得手段104（あるいはアクセス類似度解析手段501）に渡す。属性抽出手段105とハイパーリンク遷移頻度獲得手段104では、HTTP（ハイパーテキストトランスファプロトコル）に従って該URLを発行して、解析対象Webサーバとアクセスして必要な情報を獲得する。ここで、本発明の変形として、解析対象指定手段1001が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、解析対象指定手段1001は、個々のユーザーからユーザーID（あるいはユーザーIDとパスワード）を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。さらに

異なる変形として、本発明では、解析対象指定手段1001をWebサーバ上に置き、CGI（コモンゲートウェイインタフェース）スクリプトとして表現して、ユーザーのコンピュータ1002における解析対象指定をWebページのフォームを用いて行えるようにしてもよい。

【0079】[実施例7] つぎに本発明の実施例7について説明する。

【0080】図11は、本発明の実施例7を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、コンテンツ伝送手段1101により、本ハイパーテキスト解析装置のユーザーが、解析対象となるWebサーバ上のコンテンツを本ハイパーテキスト解析装置に受け渡せるようにする。本実施例では、まず、ユーザーのコンピュータ1002で、解析対象となるWebサーバ上のコンテンツを獲得する（あるいは予め解析対象となるWebサーバ上のコンテンツを持っている）。ユーザーのコンピュータ1002とコンテンツ伝送手段1101はネットワーク（例えばインターネットや電話回線）で接続されており、ユーザーのコンピュータ1002から解析対象であるWebサーバ上のコンテンツがネットワークを通じてコンテンツ伝送手段1101に渡される。コンテンツ伝送手段1101は、該Webサーバ上のコンテンツを属性抽出手段105とハイパーリンク遷移頻度獲得手段104（あるいはアクセス類似度解析手段501）に渡す。Webではコンテンツの改変が頻繁に行われることもあるが、こうすることによって、ユーザーが想定している（時点での）コンテンツを対象とした解析を確実に行うことができる。ここで、本発明の変形として、コンテンツ伝送手段1101が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、コンテンツ伝送手段1101は、個々のユーザーからユーザーID（あるいはユーザーIDとパスワード）を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。

【0081】[実施例8] つぎに本発明の実施例8について説明する。

【0082】図12は、本発明の実施例8を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、アクセス履歴情報伝送手段1201により、本ハイパーテキスト解析装置のユーザーが、解析対象となるWebサーバのアクセス履歴情報を本ハイパーテキスト解析装置に受け渡せるようにする。本実施例では、まず、ユーザーのコンピュータ1002で、解析対象となるWebサーバのアクセス履歴情報を獲得する（あるいは予め解析対象となるWebサーバのアクセス履歴情報を持っている）。ユーザーのコンピュータ1002とアクセス履歴情報伝送手段1201はネットワーク（例えばインターネットや

電話回線)で接続されており、ユーザーのコンピュータ1002から解析対象であるWebサーバのアクセス履歴情報がネットワークを通じてアクセス履歴情報伝送手段1201に渡される。アクセス履歴情報伝送手段1201は、該Webサーバのアクセス履歴情報をハイパーリンク遷移頻度獲得手段104(あるいはアクセス類似度解析手段501)に渡す。Webサーバのアクセス履歴情報は個人のプライバシーに関する情報であると考えられるため機密性が高い。そのため、Webサーバのアクセス履歴情報が該Webサーバ上でアクセス可能な形態で保持されることは希である。本実施例のように、ユーザーと本ハイパーテキスト解析装置間で直接、アクセス履歴情報を受け渡すことができれば第三者からアクセス履歴情報にアクセスされる可能性が小さくなる。ここで、本発明の変形として、アクセス履歴情報伝送手段1201が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、アクセス履歴情報伝送手段1201は、個々のユーザーからユーザーID(あるいはユーザーIDとパスワード)を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。

【0083】[実施例9] つぎに本発明の実施例9について説明する。

【0084】図13は、本発明の実施例9を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、属性指定手段1301により、本ハイパーテキスト解析装置のユーザーが、解析対象となるハイパーテキストのコンテンツから抽出する属性の組を指定することできる。ユーザーのコンピュータ1002と属性指定手段1301はネットワーク(例えばインターネットや電話回線)で接続されている。ユーザーのコンピュータ1002では、抽出する属性の組が入力され、この情報はネットワークを通じて属性指定手段1301に渡される。属性指定手段1301は、該情報を属性抽出手段105に渡す。属性抽出手段105では、指定された属性の組を解析対象として処理を行う。ここで、本発明の変形として、属性指定手段1301が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、属性指定手段1301は、個々のユーザーからユーザーID(あるいはユーザーIDとパスワード)を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。さらに異なる変形として、本発明では、属性指定手段1301をWebサーバ上に置き、CGIスクリプトとして表現して、ユーザーのコンピュータ1002における属性の組指定をWebページのフォームを用いて行えるようにしてもよい。この場合、抽出可能な属性を予めメニューとして用意しておき、該フォーム上でユーザーが選択するような形態にしておいてもよい。

【0085】[実施例10] つぎに本発明の実施例10について説明する。

【0086】図14は、本発明の実施例10を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、ユーザーのコンピュータ1002と編集手段701がネットワーク(例えばインターネットや電話回線)で接続されている。ユーザーのコンピュータ1002における編集作業はネットワークを通じて編集手段701に渡される。ここで、本発明の変形として、編集手段701が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、編集手段701は、個々のユーザーからユーザーID(あるいはユーザーIDとパスワード)を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。さらに異なる変形として、本発明では、編集手段701をWebサーバ上に置き、CGIスクリプトとして表現して、ユーザーのコンピュータ1002における編集作業をWebページのフォームを用いて行えるようにしてもよい。

【0087】[実施例11] つぎに本発明の実施例11について説明する。

【0088】図15は、本発明の実施例11を示す構成図である。尚、上記の実施例と同一部分には同一符号を付して重複する説明は省略する。本実施例は、ユーザーのコンピュータ1002と表示手段109がネットワーク(例えばインターネットや電話回線)で接続されている。表示手段109における表示結果はネットワークを通じてユーザーのコンピュータ1002に渡され、ユーザーのコンピュータ1002において表示される。ここで、本発明の変形として、表示手段109が複数のユーザーのコンピュータ1002と接続されていてもよい。この場合、表示手段109は、個々のユーザーからユーザーID(あるいはユーザーIDとパスワード)を入力してもらって、既存のデータベースアクセス制御手法を用いて、個々の処理が干渉しないような制御をする。さらに異なる変形として、本発明では、表示手段109をWebサーバ上に置き、表示結果をHTMLフォームで表現して、ユーザーのコンピュータ1002において表示されるようにしてもよい。

【0089】

【発明の効果】以上説明したように、本発明によれば、Webサイト上の任意のページ組について、ページコンテンツから抽出される様々な属性とページ間遷移頻度との相関を基にして、ページ構成の問題の原因を提示し、さらに、その改善策をも提示することができ、ハイパーテキストシステム(Webサイト)管理者は容易にページ構成を改善できる。

【図面の簡単な説明】

【図1】 本発明の実施例の構成を示すブロック図であ

る。

【図 2】 本発明の実施例の属性の一例を示す図である。

【図 3】 本発明の実施例の表示の一例を示す図である。

【図 4】 本発明の実施例の構成を示すブロック図である。

【図 5】 本発明の実施例の構成を示すブロック図である。

【図 6】 本発明の実施例の構成を示すブロック図である。

【図 7】 本発明の実施例の構成を示すブロック図である。

【図 8】 本発明の実施例の表示の一例を示す図である。

【図 9】 本発明の実施例の表示の一例を示す図である。

【図 10】 本発明の実施例の構成を示すブロック図である。

【図 11】 本発明の実施例の構成を示すブロック図である。

【図 12】 本発明の実施例の構成を示すブロック図である。

【図 13】 本発明の実施例の構成を示すブロック図である。

*

*【図 14】 本発明の実施例の構成を示すブロック図である。

【図 15】 本発明の実施例の構成を示すブロック図である。

【符号の説明】

101 Webサーバ
102 ハイパーテキスト
103 アクセス履歴情報
104 ハイパーリンク遷移頻度獲得手段
105 属性抽出手段
106 相関解析手段
107 相関データ記憶手段
108 属性解析手段
109 表示手段
401 解析対象ページ組獲得手段
501 アクセス類似度解析手段
601 解析対象ページ組獲得手段
701 編集手段
702 コンテンツ改変効果解析手段
1001 解析対象指定手段
1002 ユーザーのコンピュータ
1101 コンテンツ伝送手段
1201 アクセス履歴情報伝送手段
1301 属性指定手段

フロントページの続き

(51)Int.Cl.⁷

識別記号

F I

テーマコード(参考)

G 0 6 F 15/419

3 2 0

(72)発明者 山根 洋平

神奈川県足柄上郡中井町境430 グリーン
テクなかい 富士ゼロックス株式会社内

F ターム(参考) 5B075 KK07 ND36 NS10 PQ02 PQ14
PR03

5B082 EA00 FA11

5B089 GB04 JA22 JA37 JB02 JB14

KA13 KA14 LB14

5E501 AA02 AB15 AC25 AC35 CA02

DA02 DA17 FA14